In search of performance

van der Weijden, I.C.M.

2007

In search of performance

Research management within the Dutch public medical and health sector

ISBN-10: 90-9021269-8
ISBN-13: 978-90-9021269-2

Cover design: S. van der Ploeg, Room for ID's, Nieuwegein
Photos: Istock
Printing: PrintPartners Ipskamp B.V., Enschede

VRIJE UNIVERSITEIT

In search of performance

Research management within the Dutch public medical and health sector

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op vrijdag 19 januari 2007 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Inge Christina Maria van der Weijden

geboren te Heemstede

promotor:	prof.dr. E.C. Klasen
copromotoren:	dr. P. Groenewegen
	dr. T.C. de Gilder

Voor Kim

We think that we will have pleasure
or delight if we fulfil a certain plan,
if a certain dream comes true,
if someone we care for likes us,
if we take a wonderful trip.
This attitude is an insult to who we truly are.
We are the pleasure, we are the joy!
Why would you want your happiness
to depend on something
other than your own nature?

(vrij naar) A.H. Almaas

Thesis committee
Dr. G. van Ark (Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie)
Prof. dr. O.A.M. Fisscher (Universiteit Twente)
Prof. dr. W.H. Mager (Vrije Universiteit)
Prof. dr. A. Rip (Universiteit Twente)
Prof. dr. E.J. Ruitenberg (Vrije Universiteit)

Table of contents

CHAPTER 1: INTRODUCTION

1.1 Introduction

This thesis deals with research performance in the academic medical and health sector. Ideas about research performance are usually conceived in terms that allow for some form of comparison of a particular work with other research in the same field. Typically, these comparisons are conducted by experienced colleagues, i.e. peers. The evaluation of research, including medical and health research, by peer review is becoming increasingly problematic. It is, for instance, difficult to find non-affiliated, independent peers because of the concentration of research facilities in fewer and larger centres (King, 1987). In evaluating basic or fundamental research and strategic research[1] characterised by a high level of uncertainty compared to applied research and experimental development, it is important to recognize the various expected but also unexpected applications of this research to different areas. Medical and health research is based on the translation of biological and medical research into clinical practice, also called 'from bench to bedside' (Horton, 1999). In an increasing number of laboratories this translation can be observed literally: basic scientists and clinical scientists are working side by side, often on the same problems. Some scientists argue that in evaluating medical and health research the translational character must be taken into account, for example: *'…the scoring system used by organizations giving grants for health-related research needs to have two components: one for the scientific merits of the application, and a second for its applicability to human disease. This should receive equal recognition'*…(Horton, 1999, p. 213).

Other medical and health scientists emphasise that scientific knowledge is, after all, created to be used by clients such as doctors, social workers, patients and managers. Researchers are responsible for both knowing their target group and directly communicating their research results to the target group. For example: *'An indicator of societal relevance of research could be an article that is written in the language and medium that is used by the target group. Others are direct actions of the researcher(s) to communicate with the users*[2]*'* (Klazinga and van Everdingen, 2001, p. 35-36).

[1] The basic research category has been subdivided into 'pure or curiosity-driven research' and 'strategic research'. Pure research corresponds to the traditional notion of academic research carried out with the aim of producing new knowledge primarily for its own sake. Strategic research differs from pure or curiosity-driven research in the rationale of its support, there being at least some expectation that it will contribute background knowledge required in the development of new technologies (Irvine & Martin, 1984).

[2] This quotation was translated into English (by the author of this thesis) as literally as possible.

As stated in the quotations above, the demands for the evaluation of research have changed considerably over the past three decades. This has resulted in, what some call, a morbid growth in evaluation[3] methods, procedures and indicators. Evaluation of research and research institutions is seen as an important element in quality control of the R&D infrastructure (e.g. Rip and van der Meulen, 1995). However, some think that all this measurement has become an obsession (Harvey, Pettigrew and Ferlie, 2002). Others consider it as a new disease, the quality disease, that has affected the whole world (Vroeijensteijn, 1995). Therefore, the general objective of this thesis is to gain more insight into the evaluations organised to assess performance[4] and manage academic research in the Netherlands and their impact. It concentrates on the research work level, the research groups consisting of a head (mainly a professor) and scientific, technical, analytical and administrative staff in universities and research institutes. The focus is on medical and health research carried out in an academic setting. The medical and health sector has a long tradition of research evaluation and quality assurance. The next sections (1.2 and 1.3) briefly introduce the trends emerging in research production and evaluation. In section 1.4, health research is defined and its funding structure in the Netherlands explained. Section 1.5 addresses the research objectives and questions of this thesis. Finally, the last section, 1.6, gives an overview of this thesis.

1.2 Increased organisation and the use of research evaluations

1.2.1 Externally organised research evaluations

From the end of the 1980s[5], the government, intermediary organisations (defined as organisations or parts of organisations that have a dual alliance with both the government and the scientific

[3] The term 'research evaluation' is defined in this thesis by adopting the definition of Vroeijensteijn (1995): *'every structured activity that leads to judgement of the teaching/learning process and/or research, whether self-assessment or assessment by external experts'*. The words 'assessment' and 'evaluation' are interchangeable and used in this thesis as synonyms. There may be some difference in connotation (evaluation more than assessment seems to refer to ex-post judgements), but neither in the literature nor in practice does a strict division exist between the two (see for example Spaapen, 1995 and Vroeijensteijn, 1995).

[4] The concept of quality refers intuitively only to the scientific quality of research and to publication measures. The concept of performance was introduced in the 1970s. It is a broader concept and contains elements that are also used in the quantitative part of this thesis.

[5] For an overview of historical developments in science policy in the Netherlands and the introduction and creation of the Dutch evaluation system from the 1970s, see Rip and van der Meulen (1995) and Spaapen (1995).

community) and social groupings in the Netherlands have been putting strong emphasis on the evaluation of research. This has resulted in the development of a large number of research evaluations. From the literature, four different goals of externally organised research evaluations can be derived:

(1) To gain insight into the scientific quality of research in order to stimulate or retain (in the case of excellence) research. On the macro level, for example, the Ministry of Education, Culture and Science emphasised in the Science Budget of 1997 the view of the Dutch government on the importance of research quality in The Netherlands: …*'For the cabinet, making choices also entails choosing quality, both of institutions and individual researchers. Academic research is after all first and foremost concerned with people. The cabinet aims to retain top talent and to create the conditions for eminent research'*[2] (Ministerie van Onderwijs, Cultuur en Wetenschappen, 1996).

(2) To gain insight into the efficiency and effectiveness of public spending in order to optimise resource allocation (e.g. OECD, 1999). The budgets for scientific research in the Netherlands as a percentage of the Gross Domestic Product (GDP) have been declining, compared to the prosperity, for years. In 1987, the share in the GDP was 2.3%. In the last decade the GDP share decreased from 2.1% in 1996 to 1.89% in 2001. The Heads of State and Government of the European Union launched in 2000 a series of ambitious reforms at national and European level (the so-called Lisbon strategy). By establishing an effective internal market, by boosting research and innovation and by improving education they aimed to make the European Union "the most dynamic and competitive knowledge-based economy in the world" by 2010 (www.europa.eu.int.). A year later in Barcelona it was decided to reserve 3% of the GNP for R&D. In 2005 -half-way through the process- results were not very satisfactory. By reviewing the process Kok concluded that the implementation of reform in Member States had been quite poor (Unknown, 2004). Responsibility and commitment are the major problems. An obligatory allocation of tasks between the EU and member states must overcome this problem (Unknown, 2005) because the EU cannot boost productivity and employment if Member States do not do their part.

(3) To account to research organisations, funding agencies, government and society at large for the spending of public money. Various factors have resulted in a need for greater accountability and selectivity in the allocation of funds. In a review, King (1987) reported some of these factors, for example: an increase in the capital intensity of research; expanding objectives and opportunities, with many new fields emerging; an increase in collaborative, often

multidisciplinary, research projects which require coordination; a coalescence of basic and applied research, with much research being now of a strategic nature; economic constraints requiring choices to be made between different disciplines, fields and research proposals.

(4) To demonstrate the relevance of research for users. Governmental science policy in the Netherlands and most other Western countries includes making academic research not only more 'efficient' and 'accountable' but also more 'relevant'. As far as medical and health research is concerned, the public ultimately supports and provides money to generate improvements in patient care. The general public is also interested in the 'cultural' aspects of medical and health research, as appears from the interest in medical television programmes such as for example surgeries, the working of the brain, and information about DNA. However, Horrobin (1990) showed that if improved medical care is not delivered, support for medical research (and hence for medical journals) will dwindle and atrophy. The public, represented by patient organisations, also want to be involved in medical and research policy and management. Wagner (1999, p. 14-15), for example, reported that *'patients are no longer willing to be passive participants in the care process, but want to choose who treats them and where they are treated, and they also want to be involved in clinical and policy decision-making'*. These expectations and pressures have led to changes in criteria or indicators, methodologies and procedures traditionally used in externally organised research (including medical and health research) evaluations. For example, the need to evaluate the utility of the outcomes of research has led to the development of methods and tools for assessing the societal quality[6] of research (see 2.3).

So far the goals of externally organised research evaluations have been presented as being only important for governmental agencies, intermediary organisations, funding agencies and society at large. However, externally organised research evaluations are also important for research groups and individual scientists. For example, research grants are transformed into resources (researchers and equipment), which produce data used in knowledge claims. Papers are written and published. These papers are read and cited by the international scientific community. In this way, individual scientists and research groups accumulate credibility – credit and acknowledgement of knowledge claims – which is transformed into reputation and the chance of having subsequent proposals funded (e.g. the credibility cycle of Latour and Woolgar in Rip 1999b). Furthermore, funding agencies ask researchers to peer-review research proposals and external evaluating authorities ask researchers to participate in their external review committees to evaluate research output. Researchers are questioned in these peer review processes on the basis of their so-called

[6] The concept of the 'societal quality' of research was first introduced in Dutch discussions in the early 1990s, by for example Spaapen (1995) and van der Meulen and Rip (1995).

credibility. But, these peer review activities also increase the reputation of individual researchers. In this way, the credibility cycle for scientists is linked with credibility cycles for external authorities (including funding agencies).

1.2.2 Internally organised research evaluations

The goals of externally organised research evaluations, as described above, lead to complex evaluations at various levels, depending on organisational resource dependence. All these evaluations and resource distribution mechanisms result in a varied environment of evaluations in which we are interested. At the same time, research organisations, departments and groups have also developed and introduced internal procedures and systems themselves. These research evaluations are called internally organised research evaluations.

Various social, economic and political factors have changed the working environment for scientific research (Morris, 2002) and resulted in a need for internally organised research evaluations:

(1) Increased pressure caused by externally organised research evaluations. The number and variety of external evaluations, occurring in all phases of the scientific process (as described in 1.2.1), have increased. Internally organised research evaluations are often developed in response to the increased attention and importance of evaluations organised by actors in the environment of research groups. In general, the goals of both externally and internally organised research evaluations are similar. Improving research in order to gain a better competitive position and to obtain a leading position within the research field, e.g. van Erp Taalman Kip (1993), are stressed as the most important objectives of internally organised research evaluations.

(2) Political changes, including the new government policies for science. In the Netherlands, the focus of science policy has recently changed. Before 1998, science policy in the Netherlands focused strongly on the process of decision making by the government with respect to the development of scientific disciplines and research needed for societal areas. After 1998, the role of the government changed, giving the primary responsibility for these choices to the various research organisations (Van Steen, 2000). Today, the autonomy of research organisations (including universities) is almost complete (Meijerink, 1999).

(3) Process evaluations. Quality management in the (pharmaceutical) industry has led in the past two decades to the formulation of elaborate quality systems. Some examples of this are the rules

for Good Laboratory Practice (GLP) and Good Clinical Practice (GCP). For the production of pharmaceutical drugs, the rules for Good Manufacturing Practice (GMP) should be mentioned. A combination of these rules and standards can result in an efficient system of Total Quality Management (TQM). There is a trend to use these systems more and more in conducting scientific research in order to promote reliability and efficiency. Furthermore, these systems are sometimes required under government regulation or by customers (van der Weijden, 1999a and 1999b).

(4) To handle scientific dishonesty. The recognition of the need for systems to handle the problem of scientific misconduct has increased over recent years (e.g. Klasen and Overbeke, 2002; Koenen, 2000; Nyelenna et al., 1999). These systems may include guidelines for good scientific practice and the promotion of scientific integrity, the definition of dishonesty, procedures and bodies to prevent, detect, investigate and punish misconduct whenever it occurs, and even research into this field. In the Netherlands, the Amsterdam Medical Centre (AMC) recently developed a research code, an internal guideline ('How to act scientifically with care and integrity') for desirable behaviour (Vermeulen, 2002).

To summarise, the increasing development and involvement in externally as well as internally organised evaluations in different phases of the research process indicate that researchers and research groups are experiencing to a greater extent control over their work.

1.3 The changing context of science

As a result of the governmental cuts in public funding, as described in 1.2.1, and the governmental rethinking of the rationale for the public support of research in general, universities have reformulated their missions and management to demonstrate economic and social benefits (Morris, 2000). Various steps have been taken to exploit the results of research conducted in universities and to encourage more entrepreneurial activities (Etzkowitz, 2002; Louis et al., 1989). One way to utilise academic research in a commercial manner is to set up university spin-off companies (Meyer, 2003). Also industrial firms have increasingly attempted to develop direct connections with universities to gain access to new ideas, to encourage 'relevant' research and to promote the transfer of knowledge and technology from university centres to their own more development-oriented laboratories (Clark, 1995). These changes in universities were in keeping with broader socio-political trends, including changes in the production of scientific knowledge.

In the remaining part of this subsection some changes now unfolding in the area of knowledge production and in science systems will be introduced and the outline of discussions between scholars will be presented. These discussions will be presented because they form a relevant background for this thesis. A study that has attracted a lot of attention is that of Gibbons, Limoges, Nowotny, Schwartzman, Scott and Trow, who argued that a new kind of production of knowledge is emerging in the dynamics of science and research (Gibbons et al., 1994). This has been labelled Mode 2 knowledge production, and is emerging alongside the more traditional Mode 1. Mode 1 is discipline-based and a distinction is made between what is fundamental and what is applied. This implies an operational distinction between a theoretical core and other areas of knowledge production where theoretical insights are translated into applications. This contrasts with Mode 2 production, which does not cast research problems within a single disciplinary framework but is transdisciplinary. It is characterised by a constant flow back and forth between fundamental and applied research, between theory and the practice. In Mode 1, research groups tend towards homogeneity in terms of skill and experience. Mode 2 knowledge production is heterogeneous in terms of the skills experienced people bring to it. Mode 2 research groups are less firmly institutionalised than Mode 1 research groups. People come together in temporary working teams and networks that dissolve when a problem is solved or redefined. As compared with Mode 1, Mode 2 is more socially accountable; sensitivity to the impact of research is built in from the start. Finally, operating in Mode 2 makes all the participants more reflexive. This is because the issue on which the research is based cannot be answered in scientific and technical terms alone (Gibbons et al., 1994). Another concept describing changes in the research system is the Triple Helix Model (Etzkowitz and Leydesdorgg, 1995). These changes are expressed in terms of the relations between institutes of science, government and industry. The authors propose to model this complex system as a triple-helix DNA-like structure of university-industry-government relations (Etzkowitz and Leydesdorgg, 1995) and to present it as a methodological tool. The three strands intertwine and are cross-linked. It generates a knowledge infrastructure in terms of overlapping institutional spheres, with each taking the role of the other and with hybrid organisations emerging at the interfaces. Morris (2002) suggests that the dynamics in the model contribute to a new way of making science policy and fostering innovation. According to Etzkowitz and Leydesdorff (2000), the triple-helix thesis implies that the university can play an enhanced role in innovation in an increasingly knowledge-based society. From these two views it can be concluded that changes are occurring in the relation between science in general and its context. These changes have consequences for evaluating practices of research; the attention to societal quality is one example. The different views on changes in knowledge production and

research systems, as presented above, have also raised some criticism in the academic world. For example, Rip (1999a) argues that phrases like ‘Mode 2 knowledge production’ and ‘the triple helix’ posit a dichotomous history, which emphasises the fact that a new phase has been entered which is very different from the former pattern. He shows that when stable configurations threaten to break down, actors as well as analysts have opportunities to outline ‘a brave new world’. According to Rip (1999), this explains the fashionable interest in Mode 2 and other labels that attempt to capture a transformation that is becoming more evident. In a comment by Shinn (2002) the impact of the concepts ‘triple helix’ and ‘the new production of knowledge’ are discussed. Shinn stresses the point that changes in cognitive, artefactual and social factors are happening on such a grand scale that proof is elusive. It results in simplification and dealing in headline-grabbing metaphors where the utmost care must be taken, for example, to base concepts and conclusions on empirical studies. Especially in the medical and health-care field one wonders whether the practice of science is really changing (e.g. Blume and Geesink, 2000). The knowledge produced in the medical and health sector will, after all, be used by clients including patients. Furthermore, in order to generate improvement in patient care, users provide money to fund medical and health research (see Figure 1). The medical and health research areas are characterised by the strong relationships between researchers (in universities, non-university institutes and industrial R&D laboratories), users and funding (agencies, including pharmaceutical industries). It can be concluded that there are no grounds for arguing that there are no significant changes in the production of knowledge (Mode 2) and in the research system (Triple Helix) within the medical and health sector.

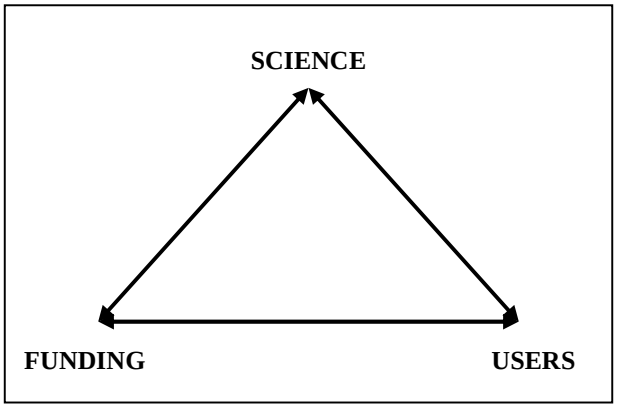

Figure 1: Relationships between science, funding (agencies) and users in the medical and health research field

1.4 Brief characterisation of Dutch medical and health research

1.4.1 Classification of medical and health research

The average life expectancy of the world population is increasing due to growing prosperity. In Western Europe also the effects of the 'babyboom' after the second world war are becoming visible. The mean age of the population is increasing which results in a corresponding increase of the demand for health care and health research and consequently puts pressure on the health care expenditure. The concept of medical and health research is broad. It covers research into the prevention, development and identification of diseases, the relief of illness and the function of the system of health care (Klasen, 2000). Two trends can be distinguished[7]: (1) scientific health research in which not only prevention and recovery from diseases are important, but also the reduction of inconvenience; (2) health care research in which the health-care system matters greatly.

The term 'health research' covers many activities, and it has been shown that it is not easy to make a straightforward classification of all these activities (Omta, 1995; Rigter, 1986). In this study, the degree to which research is related to potential patients is used to distinguish between three categories of health research: para-clinical, pre-clinical and clinical research (see Table 1). Para-clinical research emphasises both health-care research and diagnostic testing. The relationship with patients is of an advisory nature. In pre-clinical research no direct contact with patients exists. Research has a fundamental orientation and is often carried out by medical biologists and biochemists. In clinical research a direct relation exists with clinical practice. Dutch medical and health research is conducted in several academic medical centres (formalisation of cooperation in medical faculties and university hospitals) and non-university research institutes[8]. The tasks of Dutch academic medical centres are patient care, research (both biomedical, translational and clinical research), education of students (pre-doctoral), specialist medical training and additional training (both post-doctoral and post-vocational) (Buruma, 1997; NFU, 2004).

[7] The Advisory Council on Health Research, the Medical Committee of the Royal Netherlands Academy of Arts and Sciences and the Dutch Council for Medical and Health Research set up definitions of health research in October 1994. An early version of the definitions was published in 1993 by NWO Medische Wetenschappen (1993, p. 9).

[8] It should be remembered that the large variation in research quality, funding and orientation found, for example, in the USA, does not exist in the Netherlands (see Omta, 1995). Dutch universities are publicly financed and operate under the same conditions. Dutch universities as well as the Dutch non-university research institutes have, within limits, the same access to funding.

PARA-CLINICAL RESEARCH GROUPS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
Pharmacology and toxicology	Cell and developmental biology	Bioinformatics and epidemiology
Medical psychology	Genetics	Neurology
Public health	medical technology	Oncology
Social medicine	Endocrinology	Psychiatry
Environment, work and health	Immunology	Cardiovascular system
Youth and health	Microbiology	Nephrology
	Virology	Gerontology and Geriatrics
	Metabolism	Respiration
	Neurosciences	Dermatology
	Haematology	Musculoskeletal disorders
		General practice
		Gastroenterology and Hepatology

Table 1: Sub-disciplines of para-clinical, pre-clinical and clinical research groups

1.4.2 Dutch science expenditure on medical and health research

In the Netherlands, the government, charity funds and private industry fund academic medical and health research. In 1997, the Netherlands spent about 680 million Euros. This was about 10% of all national R&D expenditure (Ellenbroek, van Ark and Klasen, 2002; Klasen, 2000). Pro capita expenditure in terms of the budget for medical and health research in the Netherlands, as measured in a different manner, were relatively low (USD 55), compared to the main OECD countries. Pro capita expenditure in Germany, the UK, the USA, Denmark and Sweden was between USD 104 and USD 156 (see Table 2). A recent study by van Ark *et al.* (2005) showed that the total national expenditure for medical and health research in European countries remained at the same level (around 4‰ of the GDP) during the years 1997-2001. However, in the USA the expenditure increased steadily from 4‰ to almost 5‰ of the GDP for the same period (van Ark et al., 2005).

The allocation of resources to medical and health research in the Netherlands is complex. Responsibility for the allocation of resources is fragmented over several agencies (see Table 3). 36.5% of the total research budget flows directly to the eight medical faculties and medical clusters of the Dutch universities. The Ministry of Education, Culture and Science allocates it through the general budgeting mechanisms for universities. Governmental funds also flow from the Ministry of Education, Culture and Science and the Ministry of Health, Welfare and Sports to the Medical and Health Research Councils.

	NL	NORWAY	GERMANY	UK	USA	DENMARK	SWEDEN
Medical faculties[9]	20	33	53	6	4	20	16
Medical and health research councils[10]	4	5	3	16	46	5	5
Other governmental agencies	4	11	10	6	9	20	20
Industry	22	35	37	74	58	82	103
Charities	5	4	1	12	4	8	13
Total expenditure	55	87	104	113	120	135	156

Table 2: International comparison of pro capita medical and health research expenditure in USD. Data are from Ellenbroek, van Ark and Klasen (2002).

This agency distributes 7.0% of the budget. The remaining governmental funds account for 7.8% of the total research budget, mainly through the funding of:

– (parts of) the Royal Netherlands Academy of Arts and Sciences (KNAW) project grants;

– (parts of) the Netherlands Organization for Applied Scientific Research (TNO);

– (parts of) the National Institute of Public Health and the Environment (RIVM);

– (parts of) the Netherlands Institute for Health Services Research (NIVEL).

The charity funds which together donate about 9.1% of the funds, form an important flow of money. The pharmaceutical industry is the second most important contributor, responsible for 39.5% of the spending on health and medical R&D.

[9] Please note that in 2002 most medical faculties were co-operating to some degree with university hospitals at that time. However, this was only formalised in Leiden (Leiden University Medical Center, LUMC), Utrecht (Utrecht Medical Center, UMC) and Amsterdam (Amsterdam Medical Center, AMC).

[10] In 2001 the Health Research and Development Council (ZON) and the Medical Sciences of the Netherlands Organisation for Scientific Research (MW-NWO) were merged.

	BUDGET IN MILLION DOLLARS	% OF TOTAL
Medical faculties and academic hospitals	312	36.5
Medical and health research councils	60	7.0
Other governmental agencies	67	7.8
Industry	338	39.5
Charities	78	9.1
Total	855	100

Table 3: Dutch expenditure for medical and health research in 1997. Data are from Ellenbroek, van Ark and Klasen, 2002.

In conducting health research, academic groups in the Netherlands depend to a large extent on external financial resources. Table 3 shows that only 36.5% of the expenditure on health research is allocated by the employment organisation. Van der Weijden et al. (2002) computed, by using DAG 1998 data, the average number of PhD students and staff (in full-time equivalents, fte) per year who were paid in the period 1992–1996. They found that medical and health research, subdivided into 12 clusters, is dependent on different external financial sources (see Table 4). Dependence on the various external sources is weak in a number of clusters. For example, only 8.6% of the PhD students and staff (55.75 fte) who were working in Oncology were funded in the period 1992–1996 through the second flow of funds. The most important source of money for oncology research seems to be charity funds (third flow). These funds were responsible for funding, on average, 35.3% of the PhD positions and staff. Furthermore, in the clusters Genetics and Neuroscience a relatively high proportion of PhD students and staff (20.7% and 18.6% respectively on average) were paid via the second flow. Compared to other disciplines, research focusing on endocrinology obtained, on average, a considerable amount of money from the second and third flows of money (30.5%). Only 6.6% of the researchers were paid by pharmaceutical industries and firms (fourth flow). Finally, immunology research is mostly (58.5%) dependent on the various external resources.

	FIRST FLOW OF MONEY MEAN 49.71%	SECOND FLOW OF MONEY MEAN 12.48%	THIRD FLOW OF MONEY MEAN 24.76%	FOURTH FLOW OF MIONEY MEAN 13.06%	TOTAL NUMBER OF STAFF AND PHD STUDENTS DURING THE YEARS 1992-1996
Nutrition, digestive system, kidney and biologically active substances	60.8	10.6	20.0	8.6	539.8
Geriatrics	60.4	13.3	4.9	21.4	40.7
medical technology and biomechanics	57.4	8.9	16.9	16.8	84.5
Neurosciences	54.4	18.6	19.1	7..9	540.8
Public health	51.7	8.7	29.3	10.3	531.3
Cardiovascular disease	48.7	11.8	31.3	8.2	685
Development, Growth and differentation	47.4	15.4	24.2	13.0	261.4
Endocrinology	46.1	16.8	30.5	6.6	233.8
Genetics	43.5	20.7	24.9	10.9	328.9
Oncology	42.4	8.6	35.3	13.7	648.3
Respiratory system	42.2	3.6	26.3	27.9	52.8
Immunology	41.5	12.7	34.4	11.4	813.9

Table 4: Percentage of staff and PhD students per year who have been paid from the first, second, third and fourth flow of money[11] per discipline in the period 1992–1996[12].

1.5 Research questions

In the preceding sections, the goals, the increased interest and the use of research evaluations were described, including the changing demands for evaluation coming from different actors in the research system. The governmental cuts in the public funding of health research are reinforcing these developments. As a result, choices about research have to be made again and again. Studies such as for example (Obbink, 2000) have shown that researchers are complaining about the large number of different internally and externally organised research evaluations they are experiencing. Researchers are confronted with research evaluations at two different levels. On

[11] Financial resources are subdivided into 4 flows of money. The first flow of money is obtained from the university budget. The second flow of money is obtained from intermediary governmental research organisations such as CMHR, KNAW and the European Union (EU) research programme. The third flow of money is obtained from charitable funds. Finally, the fourth flow of money comes from industrial resources.

[12] The disciplines 'Geriatrics' and 'Respiratory system' are relative small, so it is difficult to make a statement.

the micro level, researchers are, for example, involved in the evaluation of research input. They have to submit research proposals in order to get enough research funding. Furthermore, funding agencies, in their turn, ask researchers (peers) to review proposals. On the macro level, researchers are, for instance, involved in the evaluation of research output. They are asked to collect and gather input and output data, and present it to the evaluating authority. In addition, these authorities are asking researchers (peers) to evaluate research output. Relatively little is known about the experiences of research groups with both internal and external pressures. The question arises of how research groups react to evaluation pressure and if (and in what way) this affects their behaviour. Therefore three general questions are central in this thesis.

1. **What different research evaluations have been organised to assess Dutch medical and health research?**

To gain more insight into externally organised evaluations used to assess academic medical and health research in the Netherlands, six intermediary organisations responsible for these evaluations were mapped. Intermediary organisations are organisations or parts of organisations (for example, research councils) mediating between the interests of the government on the one hand and the ongoing scientific work and its institutional arrangements on the other. The intermediaries have an important function within the environment of medical and health research groups: they are able to fund, steer, control and facilitate research. These processes could be strengthened by using the outcome of evaluations of one intermediary organisation as input into the evaluation process of other intermediaries. Intermediary organisations that primarily focus on health care, such as the Council for Public Health and Health Care (Gezondheidsraad) and the Health Care Insurance Board (College voor Zorgverzekeraars), have not been analysed. Furthermore, the evaluation practices of charities have not been considered here because the rise of their research evaluations is mainly the result of the evaluations developed and used by medical and health research councils. Consequently, the overlap between both evaluation practices is considerable.

As presented in section 1.2.2, research groups themselves recently have developed and introduced procedures and systems to assess the quality of their research. Very little is known yet about these so-called internally organised research evaluations. These evaluations could be interpreted as part of the internal management of the research organisation. It was pointed out by van der Meulen and Rip (1994, p. 41) that '*…internal management has to do with using external processes for internal policies, set up internal policies to meet external objectives and develop relations with others. The other aspect is to organize internal processes and communication optimal with*

regard to the mission of the institute'… In this thesis, research evaluations organised by universities, non-university medical research institutes, medical faculties and research groups themselves were analysed. Universities, non-university research institutes and medical faculties are also important actors in the environment of medical and health research groups. Just like intermediaries, they fund, steer, control and facilitate research. In order to answer this first research question, which is of a descriptive, comparative and explorative character, semi-structured interviews were held. Personal interviews with eight representatives of intermediary organisations and eleven personal interviews with research managers of medical faculties and non-university research institutes were held. Questions were asked about topics such as the concept of 'research quality', features of evaluation procedures (for example, goals, criteria and methods), the distinction between the scientific and societal qualities of research, and recent developments in evaluation procedures. The results of this study are discussed in Chapter 4 (externally organised medical and health research evaluations and their implications) and Chapter 5 (internally organised medical and health research evaluations and their implications).

2. **What is the impact of both externally and internally organised research evaluations?**

After outlining the evaluation practices of the most frequently used research evaluations in medical and health research, the question arises of what implications these evaluations have for the research groups concerned. A few studies have focused on the implications of evaluation outcomes. Westerheijden (1997), for example, studied the impact of the national discipline-based research evaluations co-ordinated by the Associations of Universities in the Netherlands (VSNU), on the faculties and universities concerned. He demonstrated that these evaluations have gained a place in the management of Dutch universities. It seemed that the VSNU evaluations provided managers or administrators of Dutch universities with a solid basis of legitimate arguments on which to base strategic decisions. In this thesis two types of effects are identified: (1) the effect of the evaluation procedures and methods, and (2) the effect of the evaluation outcomes. An explorative research approach was chosen because no clear definition of the possible effects was determined beforehand. Therefore, the empirical work consisted of open unstructured interviews conducted with eight representatives of intermediary organisations and eleven personal interviews with research managers of medical faculties and research institutes. Results are discussed in Chapter 4 (externally organised medical and health research evaluations and their implications) and Chapter 5 (internally organised medical and health research evaluations and their implications).

3. Do research management activities and positive views or judgements on research management enhance the performance of Dutch para-clinical, pre-clinical and clinical research groups?

The developments and changes in research evaluations of medical and health research (as sorted out in the first research question) and the implications of these evaluations (as stated in research question 2) are the stimulus for investigating whether (and what) factors really enhance research performance. Previous studies have investigated the personal characteristics of productive researchers. These include personal motivation, research training, mentors, early scholars' habits, socialisation to academic values, a network of productive colleagues, resources and substantial uninterrupted time (Cole and Cole, 1967; Pelz and Andrews, 1966). It is now clear that the research performance of groups is not only affected by personal characteristics but also by environmental ones. For a selected overview see for example, Bland and Ruffin (1992). The characteristics of a productive environment investigated in this thesis are divided into (1) research management items, a composite of internal and external control and (2) contingencies such as size of the group, time allocation and age. A couple of articles have been published on different managerial aspects that improve the performance of industrial R&D (e.g. Drongelen and Bilderbeek, 1999; Efferth, 2001 and Omta, 1995). Comparably less, but still considerable, attention has been paid to the different managerial aspects of academic research. A number of papers and studies have shown that the quality of academic research groups and their group members is affected by management activities (see Chapter 2 for a detailed description). Most of these studies concentrate on only one or a few management activities. In this study a large number of management activities have been distinguished. Internally organised research evaluations are interpreted in this study as a management activity. This choice can be supported, for example, by the study of van der Meulen and Rip (1994). They showed that internally organised research evaluation in institutes is interwoven with planning and internal communication (e.g. conferences, colloquia) and concluded that internal evaluations are one of the management instruments that enhance the quality of research. Unfortunately, empirical data that could corroborate their statement are lacking. As far as is known, empirical studies that have described the relationship between different research management activities and research performance exclude internal research evaluations from the scope of research. In order to answer this third research question, a comparative study was conducted in Dutch medical and health research groups in an academic setting. Also the impact of the disciplinary setting was taken into

account[13]. The heads of the medical and health research groups received a postal survey. They were asked to answer a total of 48 questions regarding their management activities and situation-dependent items (contingencies). Two input and five output measures were used to indicate research group performance. These measures are further explained in chapter 3.

1.6 Structure of the thesis

The next chapter is a theoretical chapter about performance indicators and research groups. In the first part, the most important evaluation methods and indicators relevant to public research that were found in the literature are outlined. The second part of the theoretical chapter discusses the literature about characteristics and activities that stimulate and maintain the performance of research groups. Chapter 3 presents a number of methodological issues. Chapters 4, 5 and 6 constitute the empirical part of this study. In Chapter 4 externally organised evaluations of Dutch medical and health research are presented and analysed (research question 1). Also the implications of these evaluation procedures and outcomes are presented (research question 2). Internally organised evaluations of Dutch medical and health research and their implications are presented and analysed in Chapter 5. The third research question is quantitatively discussed in Chapter 6. The first part of Chapter 6 describes the response rate and analyses the non-response. The second part gives an overview of the results. The relationships between research management, contingencies and performance found in Dutch medical and health research groups are presented and discussed. Finally, in Chapter 7 the results are summarised, conclusions are drawn and a general discussion of the thesis is presented.

[13] In the case of published empirical medical and health research studies, differences between several kinds of health research (para-clinical, pre-clinical and clinical health research) are frequently disregarded.

CHAPTER 2: RESEARCH GROUPS: CONTEXT AND PERFORMANCE INDICATORS

2.1 Introduction

A substantial part of academic research is carried out in research groups[14] (or also called research teams). For researchers, this research group provides the most important environment for day-to-day work and the primary focus of orientation for the research of individual scientists (Andrews, 1979b). In the past it was almost axiomatic to consider science – especially fundamental science – as the province of individuals (Stankiewicz, 1979). Over the years the prevalence of the team/group organisation in science has become widely recognised. Also, institutes and research centres were mushrooming in many parts of the world (Pelz and Andrews, 1966). Since then, most individual scientists depend on such organisational forms for funding, equipment and support. In medical and health research the necessity to: (a) share expensive research infrastructure, (b) collaborate with different disciplines and (c) share various proceedings have stimulated scientists to work together in groups.

Interestingly, Stanckiewicz (1976a) found that the prevalence of research groups, units or teams[15] (composed of scientists and supporting personnel) among Swedish academic scientists (in natural sciences and technology) was strongly related to their research field. The frequency of group membership was highest in rapidly developing fields characterised by a high degree of theoretical consensus (for example, chemistry and molecular biology). In these fields, more than 90% of scientists worked in research groups. The lowest frequencies were found in descriptive and/or theoretically less crystallised fields. In a later study, Stanckiewicz (1980) suggested that the advantage of working in a group derives from the fact that groups create the socio-psychological environment (for example, interactions in terms of communication and collaboration), which is conducive to creativity and productivity. The formation of research groups may be regarded as the means to promoting intellectual synergy by achieving a high level of interaction among scientists. Nowadays, academic research groups operate in a complex environment, consisting of

[14] Cohen, Kruse and Anbar (1982) empirically investigated the social structure of scientific research groups. They created a typology of groups as well as of leadership roles, derived from organisational and social-psychological theories. Group type or type of leadership role or both were associated with other structural features of research groups (e.g. size, disciplinary composition, whether the team had co-principal investigators and included prestigious senior scientists). It can be suggested, although this was not tested in their study, that group social structure influences team interaction in ways that can affect a group's chances of successfully reaching its research objectives.

[15] In this thesis the terms 'research groups', 'research units' and 'research teams' are used as synonyms.

scientific opportunities, constraints and resources. The environment influences the manner and direction in which scientific knowledge develops (Groenewegen, 1988). The environment of academic research groups comprises various actors, which are the key elements of the national innovation system. The research groups have interactions and exchange relations with all actors in their environment. The nature of the exchange relations between a research group and its environment varies and could change over time. From the environment, research groups draw resources (for instance, funds, legitimacy, information, instruments and know-how) in exchange for research products such as research problems, outcomes and credibility. From the environment they also exchange demands and needs that research problems may address (Zeldenrust, 1988). In order to receive resources and to define demands and needs, both actors and the research groups themselves use methods to evaluate the performance of research. In this way, each actor to some extent adds weight to the direction and quality of research conducted by the research group. The first part of this theoretical chapter gives an overview of the most important research performance measures found in the literature. In section 2.2 both qualitative (peer review process) and quantitative (number of publications and citations and more advanced bibliometric indicators) methods used in the evaluation processes of scientific research are presented and discussed. But, what factors stimulate research group performance? Some conditions for public scientific research – increased demand for accountability of public funds, increasing emphasis on both the scientific and the societal quality of research, and the redirection and reduction of funding (see Chapter 1) – increase the urgency to find answers to this question. The second part of this chapter highlights empirical contributions that help to provide insight into the capabilities of individual members as well as the environment in which the group is embedded. During 1960-1980 important contributions to research on the personal characteristics of researchers and their relationship with (both individual and group) performance were made by Pelz and Andrews (Pelz and Andrews, 1966; Andrews, 1979b). Personal motivation, confidence in one's own ideas, diversity in professional activities and skills, a moderate level of autonomy and creative ability are some examples. After 1980 this type of research was seldom continued or improved upon by other researchers. It can be argued that research organisations and research groups provide more than facilities for their members (e.g. Andrews, 1979). They also provide an environment that may either stimulate or inhibit the performance of their members. For example, Paolillo and Brown (1979) state that R&D managers should not assume they can simply hire good people and let the system run by itself. The literature strongly indicates that environmental elements of managerial control such as communication, a reward system, coordination, leadership, climate and culture, which are largely the result of the institutional setting in which the scientists are

employed, can have a substantial positive impact on the research performance of both individuals and groups. In fact, several studies suggest that environmental characteristics are the most powerful predictors of the research productivity of both researchers and research groups (e.g. Allison and Long, 1990; Long, 1978; Long and McGinnis, 1981; Thamhain and Wilemon, 1987). Compared to studies on the personal characteristics of researchers, that have been done by only a few researchers (mainly Pelz and Andrews), the environmental characteristics of researchers and groups have been studied more extensively and by a diverse set of authors. However, it should be kept in mind that elements of managerial control do not operate in research groups as isolated characteristics. *'Rather, they are like delicate threads of a whole fabric: individual, yet when interwoven, providing a strong, supportive and stimulating backdrop for the researcher'* (Bland and Ruffin, 1992; p. 387). Environmental factors of research groups that are related to individual and group performance are presented and discussed in section 2.3.1 and 2.3.2. Section 2.3.3 describes the personal characteristics of 'productive' scientists. The objective of these sections is not to give an exhaustive review of the literature but rather to highlight some issues. In both sections results of studies performed in R&D settings have been selected.

2.2 Research evaluation methods

The evaluation of research plays an important role in both science and society today. For researchers these evaluations are important, for example in competition and decisions about appointments, promotions, tenure positions, allocation of research funding and the publication of papers. Evaluations (organised either internally or externally) are not only crucial throughout a scientist's career but also in the survival of research groups, research departments and faculties, and research institutes. In this section the most important qualitative and quantitative research evaluation methods are presented and discussed. This gives some insight into the relevant dimensions of research and presents useful indicators for research group performance. Most of these performance indicators are also used by scholars in studying the relationship between managerial control, contingencies and the performance of research groups (see section 2.3.1 and 2.3.2). Also, a number of these performance indicators are used to measure the individual performance of scientists and to describe the personal characteristics of productive and effective scientists (see section 2.3.3). In addition, performance indicators were used in the empirical studies reported in this thesis (Chapters 4–6).

2.2.1 Peer review

Peer review is the predominant method for making actual evaluations of research performance in several important areas. Peer review can be defined as: …*the judgement by scientists (or other professionals identified as having the appropriate expertise) on the research (either completed or proposed) of other scientists*… (Wood, 1997, p. 9). Peer review may be confined to questions of scientific excellence or be extended to include broader socio-economic considerations. Most authors of articles on the subject assume that the purpose of peer review is quality control. According to Horrobin (1990, p.1438), in medical and health research, an additional goal of peer review is to facilitate the introduction into medicine of improvements in cure, relief and comfort. *Even in those many fields of medical research that are remote from clinical practice, the peer reviewer should always be asking the question: 'Is this a possible innovation that should be encouraged because at some time it could lead to improvements in the treatment of patients?'.*

The first precursors of the peer review process started in the scientific societies and academies of the late seventeenth century. They were crucial to the social development of the scientific journals. From the earlier practice of merely putting manuscripts into print, without competent evaluation of their content by anyone except the author, the practice of having the substance of manuscripts appraised there slowly developed. This was carried out principally before publication although sometimes after, through evaluation by institutionally assigned and ostensibly competent reviewers (Merton, 1973)[16]. At present, the referee system has been widely adopted in the academic journals of all disciplines. Most medical journals have similar practices (see Table 5). In fact, peer review has become an integral part of the social institution of science[17]. Besides for the appraisal of the quality of papers before publication in scientific journals, peer review has been introduced in many other contexts and on different aggregation levels. Peers who evaluate academic performance for career decisions, such as hiring and promotion (Creswell, 1985), are one example. Peer review has also been introduced into the evaluation of research output.

[16] For a detailed (historical) analysis of the introduction and development of peer review in science, see, for example, the chapter 'Institutionalized patterns of evaluation in science' by Merton and Zuckerman, published in The Sociology of Science: theoretical and empirical investigations.

- One to three reviewers
- Reviewers selected from senior academics
- Reviewers know author's name and institutions
- Reviewers do not sign their reviews
- Authors do not know reviewers' names and institutions (blind review)
- Simple, general instructions to reviewers
- Ask reviewers to provide: comments for authors, comments for editors, grades on individual components and overall quality of the manuscript

Table 5: Conventional peer review in medical journals, after Fletcher and Fletcher (1999)

The evaluation of research programmes of universities and institutes, research schools and disciplinary evaluation practices (see for an example of evaluation of health research in the Netherlands in 1988, 1994 and 1999) are based on peer review. Furthermore, the peer review process is common practice in research councils and other funding agencies (e.g. (Bazeley, 1998; Cole, Rubin and Cole, 1978; Wood, 1997). Peers evaluate research proposals and have positions in advisory committees. According to Rip (1999b), research councils became legitimate to scientists in this way. Effects of this can be seen on reward and reputation systems. Grants received from the national research councils and international funding organisations became indicators of performance and credit. For example, in job applications, tenure decisions and even in the allocation process of external funds (through peer review of the research proposals), it is important to show that grants have been awarded.

New forms of peer review within medical and health research

The most important recent change in the publication processes, electronic publication, is affecting peer review of articles (Rennie, 1999). One example is *open peer review*, where the names of reviewers (in contrast with formal peer review) are known to the author. In medical and health sciences, the British Medical Journal has, after completing a randomised controlled trial (Van Rooyen, 1999), introduced open peer review and is experimenting with an open electronic system (for more detailed information, see http://bmj.bmjjournals.com/advice/peer_review.shtml).

[17] Since 1990, even international congresses have been organised to discuss the process of medical peer review, the most recent insights into what has been achieved by peer review and how it might be improved (Goldbeck-Wood, 1999).

Another type of peer review is a *refereeing structure* with review comments made public and all comments, responses, and changes becoming part of the final publication. In the medical and health sector a system of this kind is already in practice in the journal, Behavioral & Brain Sciences, where, after formal review, articles are circulated to as many as one hundred potential commentators, across specialities and around the world. Then, each article is co-published with the 20 to 30 (accepted) peer commentaries it elicits, plus the author's response to the commentaries (Harnad, 1996). The last type mentioned here is *post-publication*. The Medical Journal of Australia, for instance, performed an Internet peer review study of electronically published articles (Bingham, 1999). After electronic publication the authors could revise the publication in response to comments from readers. Afterwards, the articles were published in print.

Problems with peer review

In an ideal world, there would seem to be only advantages involved in any well-run, smoothly functioning peer review system. Decision makers (e.g. editors, funding agencies) feel more comfortable in their decisions when they are informed by expert opinions. Peers appreciate being recognised as experts. Scientists who are evaluated (e.g. applicants, authors) realise that the only hurdles worth jumping are those they respect: the hurdle erected by experts (Rennie, 1999). Over the years a lot of articles have been published that criticise (by giving their opinion and/or by collecting empirical data) the use of peer review in assessing research performance (e.g. Horrobin, 1990; Moxham and Anderson, 1992). Reviewers may, for instance, be partially biased, jealous, ignorant, incompetent, malicious, corrupt or incapacitated by conflicts of interest (Rennie, 1999). Of the many criticisms directed at peer review prominent ones are highlighted in this section.

(1) Selection of peers on the basis of their point of view

Peers need to be specialists in the relevant research field, but peers who are too closely involved with the subject may be influenced by jealousy or cronyism (e.g. Fuhrer and Grabois, 1985; Wenneras and Wold, 1997), whereas more distant referees may not have the required expertise. The use of international referees is frequently proposed as a way to reduce conflict of interest and jealousy. However, Wessely (1998) found that they tend to produce more favourable and less rigorous evaluations.

(2) Biases in the peer review process

In the peer review process several examples of bias can be found. Bias against personal characteristics of scientists, such as sex (female scientists) and research experience (young

scientists), is well known. Cognitive bias, when a referee gives a negative review because the theoretical approach or methodological approach of the author is not the same as that of the referee, is described by several authors including Cole (1998) and Wood (1997). Institutional bias, when a referee gives a negative review because the researcher works in an institute that is less known to the referee (Wood, 1997), is another example of bias in the peer review process. Finally, the Matthew effect in science is probably the best known example of the consequences of bias in psychological processes. It was originally construed in terms of '*the enhancement of the position of already eminent scientists who are given disproportionate credit in cases of collaboration or of independent multiple discoveries'* (Merton, 1973, p. 458). The Matthew effect[18] may heighten the visibility of contributions by more highly ranking scholars. For example, they are more likely to have their papers or proposals accepted, because their names are recognised by peers, than authors with lower academic qualifications.

(3) Low reliability of peer review ratings

The extent to which independent reviewers (of the same proposal/paper, etc.) agree on the quality of a paper is one indicator of peer review reliability. However, studies have shown a low level of consensus or agreement of assessor scores among peers. For example, Cicchetti (1991) reported in a review about agreement between peers in academic journals a low reviewer agreement in both medicine and behavioural sciences. Also, Rothwell and Martyn (2000) showed that agreement between independent reviewers on whether manuscripts in the field of clinical neuroscience should be published, or their priority for publication, was low. In fact, agreement was not greater than what would have been expected by chance alone.

(4) Time-consuming and costly process

Researchers spend much time on writing and reviewing research papers and grant applications. The resource input into the review process, both in terms of administrative costs and of scientists' time, is considerable but, according to King (1987), usually ignored. Not only awarding bodies but also editors of scientific journals encounter difficulties in persuading scientists to give up time to review grant applications or papers (e.g. Cole, 1998; Wessely, 1998). The Australian Research Council, for example, now makes agreements to undertake reviewing duties a condition for awarding a grant (Wessely, 1998). Another example is the former Dutch medical research council MW-NWO. In order to attract more reviewers, this council carried out an experiment in 2001,where they paid international researchers who were willing to review a research grant proposal. However, the results showed that the (extra) amount of money invested did not

[18] In the literature there is some dispute about the amount of variance explained by the Matthew effect (e.g. Cole, 1998; Daniel, 1998).

compensate for the increased number and quality of review reports (MWO-MW, 2000). After internal evaluation of this experiment, it was decided not to continue paying reviewers.

(5) Lack of transparency

Some authors argue that the peer review process has the characteristics of a black box. It does produce judgements, but one does not know how they come about. In order to get some insight into the thought processes of peers when they pass their judgements, Sonnert (1995) explored the criteria by which (in this case US biology) scientists evaluate the performance of their colleagues. He found, among a variety of productivity measures, three significant predictors – annual productivity rate, solo-authored publications and graduate school expertise – that together explained 59% of the variance in the judgements of biologists.

In response to the criticism many alternatives and additions to peer review have been suggested for measuring research performance. The most common ones are quantitative indicators based on the use of characteristics of scientific publications that in general publications are considered the most important output of scientific research. As Chubin and Hackett (1990; p. 83) stated:

…*'Publication is the lifeblood of science, conveying the symbolic nutrients of new theories, research findings, credit and critical scrutiny'*…

2.2.2 Quantitative methods

In this section the commonest quantitative measures and indicators of research performance[19] found in the literature are discussed: publication counts, impact factors, citation counts and more advanced bibliometric methods.

2.2.2.1 Number of publications

The number of publications is considered to be an indicator of the quantity of scientific output. It is obtained by simply counting the number of publications or by counting the number of publications while using a weighted scale for each type of publication (Creswell, 1985). It may

[19] Indicators other than described in this section can measure research performance. Examples are the number of research grants obtained, the number of appointments to editorial boards, the number of improved processes, the number of new products and the number of new analytical methods (e.g. Creswell, 1985). In medical and health sciences also other experimental methods based on the evaluation of scientific

include journal articles, (chapters in) books, papers presented at professional meetings and book reviews (e.g. Patrick and Stanley, 1996). Most scientists believe that the quantity of output is not the equivalent of output quality. According to Smith and Fiedler (1971), the main problem in using publication counts as a measure of research performance is that they may give equal credit to poorly conceived papers appearing in poorly edited journals and to well-written papers in high-quality journals.

2.2.2.2 Number of citations

Until 1961, there seemed to be no practical way of measuring the impact of publications. Cole and Cole (1973) reported that the problem of assessing the quality of scientific publications has long been a major impediment to progress in the sociology of science. The invention of the Science Citation Index (SCI)[20] in 1961 by Eugene Garfield and the publication of the first volume of SCI in 1963 provided, for the first time, a tool for measuring the significance of the contributions of individual scientists[21]. The Citation Indices (CI) claim to cover the most important 'leading' international journals and serials with a well-functioning referee system. More 'peripheral' journals, often national in scope, are usually not covered by the CI. By using references in publications, researchers show how they have built on previous work by others. The quality of a publication and its influence on academic knowledge are reflected in the number of times others cite it (Cole and Zuckerman, 1984). In this way, citations seem to have universal quality because any article can be compared with any other, independently of the subjects involved (e.g. Egghe and Rousseau, 1990; Wouters, 1999).

Problems with the use of citations

The use of citations, in the years after the construction of the SCI as a science performance indicator, ran into some problems regarding the gathering and handling of data. A distinction was

papers are developed, presented and discussed by scholars; see, for instance, the study of Ugolini, Parodi and Sant (1997) in the field of oncology.

[20] The SCI claims to cover the central and most important journals in the natural and life sciences. It fully indexes 5900 major journals across 150 scientific disciplines. The Social Science Citation Index (SSCI, a multidisciplinary index of 1725 journals across 50 social science disciplines), the Arts and Humanities Citation Index (AHCI, a multidisciplinary index of 1144 of the world's leading arts and humanities journals), the Materials Science Citation Index (MSCI) and the Neuroscience Citation Index (NCI)are related indices published by the Institute for Scientific Information (ISI).

[21] For a detailed analysis of the creation and building of the SCI, see the thesis of Wouters (1999), chapters 2 and 3.

made between technical and methodological problems. Technical difficulties created by the production process of citations include:

- deficiency or errors in references, caused by spelling mistakes (estimated at 10% by Moed, Burger and van Raan, 1983), leading to non-existent texts;
- regular occurrence of identical names for different entities;
- incompleteness of the ISI database. See, for example, Lewison (2002), who studied this problem in the medical field; and
- self-citations[22].

These problems have been frequently noted in the literature, e.g. (Egghe and Rousseau, 990; King, 1987; Moed et al., 1983; van Raan, 1999c). In addition, some methodological problems have been experienced, e.g. Cole and Cole, 1973; Egghe and Rousseau, 1990; Smith, 1981; van Raan, 1999c):

- (sub)field dependence: the chance of being cited varies greatly among scientific specialties or disciplines, depending on the size (number of journals covered), publication culture (importance of journals as a communication medium)[23] and citing culture of the field concerned (number of citations to a publication)[24];
- time delays for citations: the amount of time it will take to elicit citations varies greatly among (sub)fields[25];
- critical citations: citations may refer to papers that are being criticised and rejected rather than utilised;
- dominance of English as the scientific language.

Besides these technical and methodological problems, also some concern about the impact of citation counts has been raised, as is illustrated by a quote from Cole and Zuckerman (1984, p.240): *'Does a work cited an average of 55 times over 12 years have a greater impact than a work cited an average of 32 times over 12 years? Or, have two scientists who have each been cited 50 times in a given year had the same impact as one who has published 25 papers, each of which has been cited twice, and the other has also published 25 papers but has two papers which received 25 citations each, while the others 23 received none at all'?*

[22] According to Moed et al. (1983) about 10% of all citations appear to be self-citations.
[23] According to van Leeuwen (1996) books, proceedings and reports play an important role in the communication of results in application- and problem-oriented scientific fields .
[24] Van Raan (1999c) compared an average publication in biology and an average publication in mathematics. He found that a biological publication received an average of 20 citations while an average mathematical publication received only 5 citations.

2.2.2.3 Bibliometric methods

Over the years, more sophisticated bibliometric indicators have been developed, constructed and used in research evaluations. In fact, a new profession has been created, *scientometrics*, with a core journal and an international conference. The more sophisticated indicators are too numerous to discuss in detail individually. According to Wouters (1999, p. 115), there is a pattern in the construction of these indicators: *'First, all are built on the basis of varying combinations of the signs reference and citation. The way these two dimensions (the citing and the cited) are combined determines important characteristics of the resulting indicator network. Second, they all aim to represent reality in a more reliable way than competing indicators, or qualitative descriptions. Third, they build upon one another'.*

Impact factor

The impact factor (IF), introduced by Eugene Garfield and regularly published in the Journal Citation Report, is a purely citation-based measure for the significance and performance of journals. The impact factor is calculated from the number of citations to a scientific journal in a certain period divided by the number of publications in that journal during the same period (mostly 2 years). According to Glänzel and Moed (2002), it is the most popular bibliometric product used in bibliometrics as well as outside the scientific community[26]. Although the use of journal citation impact factors as a performance indicator has been practised for many years, it has not been without criticism. See, for instance, the critical letters and reviews of Boor (1982, p. 976): *'…recognize characteristics of the citation impact factor that render it of dubious validity and potentially deleterious to the research and publication enterprises*'; Edwards, van Steirteghem and Richardson (1993, p. 983):*'we have discovered a major inconsistency in the compilation of citation data for journals that are included in Current Contents, but not covered in the Science Citation Index'*; Hansson (1995, p. 906): *'impact factor as a misleading tool: there is virtually no correlation between the citation frequency and IF of the journal in which it is published*; Moed and van Leeuwen (1996, p. 186): *'…the IFs of many journals included in the Science Citation Index (SCI) are inaccurate because of an inappropriate definition of citable documents'* and Smith (1998, p. 1079): *'…impact factors are meaningless as a measure of the*

[25] Van Raan (Van Raan, 1999c) found that publications in more fundamental sciences (e.g. natural and medical fields) cite recently published work (a maximum of three years after publication). In applied scientific (sub)fields citations are made to much older publications.

[26] For a (historical) overview of the most important attempts to improve or to complement the impact factor, see Glänzel and Moed (Glänzel & Moed, 2002).

performance of individual scientists or research groups for the simple reason that there is little correlation between the number of times that individual articles may be cited and the impact factor of a journal. This is because impact factors depend on a few articles that are highly cited'. The impact factor and other bibliometric indicators are mainly used to assess the scientific quality of research. Lewison (2002) suggested that, for example, journal-specific indicators can also be used to assess some societal aspects of (in his case medical) research. For instance, the frequency with which journals are cited in clinical guidelines, in patents and in newspapers can be measures that can be used to analyse the societal aspects of research quality.

Two-dimensional indicators

In the above section only one-dimensional quantitative research performance indicators are discussed, based on direct counts of specific bibliographic items (e.g. publications) or particular data elements in these items such as citations (Van Raan and Tijssen, 1993). Box 1 and box 2 introduce two-dimensional indicators constructed from the co-occurrence of specific information elements: co-citation and co-word analysis. With the help of special data analytical techniques[27] based on matrix algebra, it is possible to convert the information in such a co-occurrence matrix into a spatial configuration of elements (keywords or references) in a two-dimensional space. In this way a map is made of a scientific field. These maps can be interpreted by scientists, science managers and science policy officials, see, for instance, Noyons (1999) and Wouters (1999). Despite the promising introduction of bibliometric maps of science in a science policy context in the 1970s, they have not been very successful yet. According to Noyons (2001), it seems, however, that only recently they are becoming acknowledged as useful tools, mainly by virtue of the developments and integration of hypertext and graphical interfaces, which makes the strength of such navigation tools more visible.

Box 1: Co-citation analysis

The first two-dimensional technique is based on the number of times citations are mentioned together in publications. The co-citation technique was introduced in 1973 (Small, 1973). The co-citation frequency is the number of times a certain pair of cited articles is cited together in the same papers[1]. Co-citation analysis is, in fact, an attempt to identify so-called 'high-density areas' in a citation network by clustering highly co-cited documents and it thereby indicates the existence of 'research fronts'[1] (Braam, Moed and van Raan, 1991). However, the validity, the interpretation of results and the practical usefulness of co-citation analyses for science policy purposes are often subjects of debate (e.g. Hicks, 1987; King, 1987; Rip, 1988).

Box 2: Co-word analysis

The second two-dimensional technique, co-word analysis, is not based on citations and references but on the number of times keywords mentioned together in publications. In 1983, co-word analysis was introduced to a wider audience by Callon and his colleagues (1983). In short, for each scientific field all keywords of publications can be collected, and for each keyword in the compiled set one may analyse how many times each keyword occurs together with any other keyword in the publications involved (Callon, 1983). Examples of bibliometric maps based on co-word analysis can be found in van Raan and Tijssen (1993: map of neural network research) and Noyons, Buter and van Raan (1999: map of neurosciences). Maps of science can be seen as tools for searching, identifying and analysing structures of scientific activities as reflected by publications. They may point to emerging fields of science and emerging new activities, and they offer insight into the position of countries, research organisations or institutes in a field of science (van Raan and Tijssen, 1993). Just like all measures or techniques, co-word analysis has its limits and potentials. According to van Raan (1999a), an advantage is that, when maps are made for a series of years, it becomes possible to observe trends and changes in the structure of fields. According to Hinze (1997), the most important disadvantage is the time delay caused by the slow actualisation of thesauri and classification schemes.

[27] For details about the sophisticated combination of clustering techniques and multi-dimensional scaling used in these two-dimensional bibliometric techniques, see the studies of Peters and van Raan (1993a; 1993b).

2.2.2.4 Application of bibliometric methods in the research evaluation process

In spite of the criticism of peer review processes, as discussed in section 2.3.1, peer review is still the most accepted way to evaluate research. Compared to some decades ago, the main difference is that peer review is no more the one and only way to evaluate research. Increasingly, quantitative methods and indicators allow the measurement of the response of the international scientific community to the published work (expressed in references in scientific literature) in the evaluation processes of science. Currently, bibliometric indicators are frequently used – often in combination with peer review – not only in externally organised research evaluations (see box 3 for some examples in the Netherlands) but also by scientists themselves. The impact factor is the most commonly used bibliometric indicator because it contributes to solving problems by comparing the output of the various scientific (sub)fields. In section 2.3.2.2 it is shown that scientific (sub)fields differ in publication and citation patterns. Therefore, it is usually not meaningful to compare the raw impact of publications in one field with those in another field (Nederhof, van Leeuwen and Visser, 1998a). In solving this problem, two international benchmark indicators were developed by van Raan (1996): (1) the mean journal citation score based on all publications of the group or institute evaluated), (2) the mean field citation score based on all publications in SCI journals in one field. In order to get information about the impact of research publications, the mean impact scores per publication are compared with the benchmark indicators. Furthermore, Noyons and Moed (1999) recently demonstrated that the combination of mapping and citation analysis can be a powerful tool in the evaluation of research activities. This combined approach was applied to support a governmental audit of research activities of the Inter-university Centre for Micro-Electronics in Leuven, Belgium. On the basis of the comments of an international panel of experts in micro-electronics, the method was discussed in detail. The panel concluded that the method provided a detailed and useful picture of the position of the institute from an international perspective. Moreover, they found that the results of each of the two parts had added value for the other one.

AGGREGATION LEVEL	EXAMPLE	EVALUATION GOAL
Individual researcher	None	
Research departments or research groups	Bibliometric analysis of economic research groups (Nederhof and van Raan, 1993)	Research performance of research units in economics has been evaluated by simultaneous efforts of peers and bibliometricians
Research institute	Bibliometric analysis of the Cardio-vascular Research Institute Maastricht (Nederhof and Visser, 1997)	Bibliometric analysis of the research institute CARIM (part of the medical faculty of Maastricht University) organised by CARIM
Faculty	Bibliometric analysis of the Faculty of Medicine of the University of Groningen (Nederhof, van Leeuwen and Visser, 1998a)	Bibliometric analysis of the Faculty of Medicine (Universitiy of Groningen), as part of the internal evaluation of research output
UMC	Bibliometric analysis of eight UMCs (NFU, 2004)	External goals: transparency and justification to Dutch society and funding agencies. Internal (UMC) goals: results can be used to make policy and improve research
Research school	Bibliometric analysis of the Netherlands School of Primary Care Research (Nederhof, van Leeuwen and Visser, 1998b)	External research evaluation of research school CaRe based on bibliometric indicators only
Research field	Quantitative assessment of Dutch soil biological research (NRLO, 1997)	Explorative study of the use of bibliometric indicators in strength-weakness analyses of soil biological research
	Bibliometric profile of Dutch physical research (van Leeuwen, 1996)	Evaluation of Dutch physical research. Bibliometric profile is used in combination with a report of the peer review committee.

Box 3: Examples of bibliometric methods in Dutch (externally organised) output evaluations

2.2.3 Societal performance indicators

Over the past years externally organised research evaluations have focused on indicators measuring the *scientific quality* of research, including the role of bibliometric analysis on scientific impact (see sections 2.3.1–2.3.2.4). The scientific quality of research is based on the view of Merton (1973), who argues that the contribution to the growth of the scientific knowledge

is the final objective of science. In addition to scientific quality, the *societal quality*[28] of research is also important. It refers to the increased pressure on researchers to demonstrate user relevance, the relationship to industry or the utility of the research, e.g. van der Meulen and Rip (2000). In medical and health research the final objective is to improve the health outcome for each individual. Therefore, societal quality can be divided into: (1) relevance for health care providers and the process of health care delivery (KNAW, 2002), (2) relevance for policy makers and the process of designing, implementing and monitoring policy decisions (Council for Medical Sciences, 2002) and (3) the relationship to the pharmaceutical industry. Besides societal quality, the *societal impact* of medical and health research – which can be described as the implementation of research output by executive professionals and the policy field – is also an important element to evaluate. It addresses the systematic and planned process of integrating new research findings or valuable procedures and techniques within normal practice routines.

Societal indicators in externally organised research input evaluations

In selecting research projects there is traditionally a peer evaluation of scientific quality. In society-inspired research programmes, an important criterion in the assessment of research proposals is their merit in relation to the defined targets of the particular programme. To obtain research grants, proposals have to score above a certain threshold on both scientific and societal indicators.

Societal indicators in externally organised research output evaluations

Over the past few years, some experimental evaluations have been carried out which tried to develop methodologies to evaluate the societal quality and impact of research output. Most of these developed and tested methodologies are quantitative evaluation tools (see Table 6) that can be used in qualitative peer practices such as self-evaluations and external audits. The criteria and indicators used in this kind of evaluations always depend on the mission of the group or institute evaluated (Council for Medical Sciences, 2002). Therefore, evaluation outcomes must always be interpreted in relation to the research mission.

[28] See for example van der Meulen and Rip (2000) who identified (in a case study approach) dimensions of societal quality in environmental as well as in neuromuscular research and indicators linked to them. They found that the expectation that the research would contribute to socio-economic developments (relevance), the interaction with (possible) users and the actual use of results were the dimensions mainly used.

- Societal quality research profile (SQRP) in health research (Spaapen, 1995)
- Research embedment and performance profile (REPP) in agricultural sciences (Wamelink and Spaapen, 1999) and pharmaceutical research (Callon, Law and Rip, 1986; Dijkstelbloem, Spaapen and Wamelink, 2002)
- Stakeholder analysis of agricultural sciences (Wamelink and Spaapen, 1999) and pharmaceutical research (Dijstelbloem, Spaapen and Wamelink, 2002)
- End-user relevance in agricultural and biological research organizations (Lyall *et al*., 2004)
- Methodology for the (ex-post) evaluation of the societal impact of applied health research (Council for Medical Sciences, 2002) in which communication to health professional workers, quality improvements, implementation activities and the relevance of health care for policy are evaluated.

Table 6: Examples of recently developed ex-post societal quality methods and indicators

Concluding remarks: The use of multiple indicators in research evaluations

In the preceding sections the most important qualitative (peer review process) and quantitative (number of publications and citations and more advanced bibliometric indicators) methods used in evaluation processes of scientific research are presented and discussed[29]. Of all these different methods, the qualitative peer review process is still the most accepted one. According to Chubin and Hackett (1990), the refereeing effort is even seen as *'a good citizen chore that comes with membership in the scientific community'*. In contrast, quantification and weighting of research output have been most difficult, problematic and debated tasks for years, e.g. Cole and Cole (1971) and Endler (1978). According to Sonnert (1995), a major strength of the quantitative methods is their reliability. Their weakness lies in the area of validity: do they really measure what they intend to measure? For instance, researchers may also tend to cite what they consider to be poor or flawed studies to contrast them with their own work. Finally, with regard to the interpretation of results of bibliometric studies, it should be clear that the results of bibliometric

[29] It has been noted in the literature that performance indicators are inter-correlated. Cole and Zuckerman (1984), for example, reported positive correlations between citations and publication counts, ranging from $r = 0.50$ to $r = 0.75$. Citation counts also correlate strongly with other performance indicators, such as employment in a prestigious university, listing in important bibliographies of scientists, and receiving scientific awards (for instance Nobel Prize winners and fellowships awarded (Creager, 1967 in Smith and Fiedler, 1971). Furthermore, citations correlate positively with peer review ratings. Cole and Cole (1971) found that visibility with peers correlates positively with publishing productivity rates ($r = 0.56$) and citation counts ($r = 0.68$). No research has been done yet on the correlation between different societal indicators because this field is relatively new and still developing.

studies could not be interpreted properly without background knowledge. Some knowledge of specific internal circumstances and the international subfield is required (van Leeuwen, 1996). It is currently a common belief, especially among bibliometric experts, that the combination of advanced bibliometric methods and peer review is the best approach in evaluating research, e.g. van Raan (1996). Furthermore, bibliometric experts emphasise that bibliometric indicators are not meant to replace peer evaluation, e.g. Nederhof and Visser (1997), van Leeuwen (1996) and van Raan (1999c), but: *'…they can offer crucial information about research performance that can be seen as complementary to peer opinion'…* (van Leeuwen, 1996, p. 3).

The research evaluation methods as presented and discussed in the preceding section present a brief overview of the environmental complexity of research groups.

2.3 Environmental and personal characteristics and research performance

Factors in the environment of research groups (both in an organisational and social context) are divided in this section into contingencies (2.3.1) and managerial control factors (2.3.2). Each of the environmental factors is discussed separately. However, it should be remembered that environmental factors do not operate in research groups as isolated characteristics. *'Rather, they are like delicate threads of a whole fabric: individual, yet when interwoven, providing a strong, supportive and stimulating backdrop for the researcher'* (Bland and Ruffin, 1992, p. 387). 2.3.3 finally presents important contributions to research on the personal characteristics of researchers and their relationship with both individual and group performance.

2.3.1 Contingencies and performance

In the literature group size, age, resources and research goals are the most important contingencies that are related to research performance. In this section an overview of the literature is presented.

2.3.1.1 Group size and performance

There have been several pioneering studies on the effects of group size on research performance, for example Wallmark and Sellerberg (1966), Blume and Sinclair (1973a; 1973b) and Wallmark *et al*. (1973). Wallmark and colleagues (1966, 1973) investigated the relationship between the size of 60 research teams in applied physics and their performance (measured as the number of citations that the team's papers received divided by the number of citations, excluding self-citations). They found that: (1) there is a positive effect of group size on performance; (2) there is no optimum team size; (3) the improvement in performance is exponential with the size of the research team, and (4) other factors such as material resources, selection of productive group members and personal characteristics could contribute to research performance of groups. Blume and Sinclair (1973a, 1973b) investigated the relationship between group size and effectiveness (measured as the number of articles) in a large sample of British academic chemists. They concluded that neither critical nor optimum size effects exist. They also found that the relationship between size and productivity varied considerably among areas of the chemistry discipline. They speculated on the multiplicity of skills required for some types of research and on the degree of mechanisation and typification of the research procedure influencing this relationship. However, in studying 200 Swedish academic groups in natural and technology sciences, Stanckiewicz (1980) found that group size affected both the quality and quantity of research output. The relationship was curvilinear with the optimum corresponding to the average size of 6–7 scientists. More recently, for example, Jordan, Meador and Walter (1988), who studied economic departments, Pineau and Levy-Leboyer (1983), who studied medical laboratories, and Baird (1986), who studied chemistry, history and psychology, showed that research performance increases with the increasing size of the research group. These findings suggest that, as one would expect, there are, for example, more opportunities for contact, stimulation, and resources in larger research groups (see also the review by Bland and Ruffin, published in 1992).

2.3.1.2 Age and performance

The first studies investigating the effect of group age on the performance of research groups were published by Wells (1962) and Wells and Pelz (1966). They studied in 83 industrial and governmental research groups the relationship between the age of groups and their performance.

Age was measured as the average number of years the members belonged to the group. Performance was measured as the rating of the scientists' scientific contribution and their overall usefulness to the organisation. They found that (1) the general scientific contribution of the groups tended to decline with increasing group age and (2) the group's overall usefulness to the organisation tended to increase during the first four to five years, after which it declined. Wells and Pelz (1966) found that these effects could be linked to decreasing cohesion and competitiveness and to increasing specialisation in 'ageing' groups. Smith (1971) found similar results by studying 52 R&D groups.

A few scholars studied age as the amount of research experience in relation to performance. Dill (1985), for instance, suggested that the leader's professional expertise as a scientist significantly affected the unit's productivity. Dill also contended that the professional experience of leaders enables them to influence members' knowledge and values, to facilitate contacts and networks, to attract other competent researchers, to help colleagues, and so on. Stanckiewicz (1980) found that the level of a leader's R&D experience bears a fairly strong, largely linear relationship to the index of recognition of the research groups, but seems to be unrelated to the groups' quantitative performance. Recently, Omta (1995) found that the research and management experience of research leaders correlates inversely with the annual growth rates of medical research groups.

2.3.1.3 Age, size and performance

Knorr *et al*. (1979) studied, with data drawn from the International Comparative Study on Organizational and Performance of Research Units, the relation between different organisation characteristics and the productivity of academic (natural science and technological science) research groups. They showed that, in addition to individual productivity, three contingency factors positively influence the published output of research groups. These are:

(a) *size* of the research group, measured as the average number of man-years of scientists;
(b) *age* of the research group, measured as the number of years a group has existed formally under its present name and goal structure;
(c) *scientific exchange* maintained by the research group.

By using the same data, Stanckiewicz (1979) found that given a high level of cohesiveness, the effectiveness of research tends to increase with group size. This result is in line with other studies, including those of Blume and Sinclair (1973a; 1973b) and Wallmark *et al*. (1973). But, when cohesiveness is low, there seems to be a strong decline in performance for research groups larger

than 7 scientific members. Stanckiewicz (1979; 1980) also reported that the relationship between size and output per scientist is strong within groups headed by leaders with long research experience (more than 14 years), but it is weak within groups headed by younger researchers. In fact, within groups headed by younger leaders, there is a negative relationship between productivity and group size. Finally, it was found that groups characterised by high levels of cohesiveness and/or directed by experienced leaders, group age has little effect on research performance. In contrast, within groups characterised by low cohesiveness and/or directed by highly involved leaders, the importance of group age is considerable (see Stanckiewicz, 1979).

2.3.1.4 Resources and performance

Resources to accomplish a task or to achieve a goal are essential for any activity, including research. Stolte-Heiskanen (1979) investigated the impact of *material resources* (such as workspace and equipment), *human resources* (such as size of the unit, contacts of the unit, competence of the unit head, unit scientists and technicians), and *information resources* (such as international library services and science information service) on administrative and social effectiveness of research units. The satisfaction of the unit's members with these resources was defined as subjective resources. Data were collected in 1222 research units from six countries. The relationships of material and information resources to effectiveness was generally very weak. On the whole, human resources seem to play a more important role in the effectiveness of research units. The possibilities for interaction with other scientists and the competence of the unit head are the most important (objective) human resources. The satisfaction with human resources explains most of the variance in the unit's performance (about 5% of the variance in administrative effectiveness).

Pineau and Levy-Leboyer (1983) highlighted the importance of *support staff*. They found that, among 155 medical teams, the least productive teams had either no or only a few full-time laboratory technicians, and the most productive teams had ten or more full-time technicians. Culpepper and Franks (1984) studied the common major impediments to research reported by 353 family medicine university units. They found that lack of time (78%) and lack of funding (61%) seemed to be the two most important impediments to research. Other studies, as for example Creswell (1985), Meltzer (1956) and Pruthi *et al*. (1993) also reported the inadequate availability of *money* as a barrier to performing research. Finally, Stanckiewicz (1980) studied the relation between the *amount of time* a research leader (n = 200) spends with the group and

performance and found that performance improved as the percentage of time allocated to the group increased up to 35%. If more than 35% of the leader's time is allocated to the group, the performance stabilised or even decreased.

2.3.1.5 Research emphasis and performance

In 1986, Baird showed that departments with clear, dominant goals of research were more productive than departments that emphasised the training of practitioners. Similarly, other departments that place high priority on practitioner training and service, such as family medicine and nursing, generally have low research productivity (Perkoff, 1986). Omta (1995) found that scientific staff of pre-clinical and para-clinical units can spend twice as much time on research as staff in clinical units. Interestingly, this extra time was not translated to higher research and user effectiveness. The researchers in clinical units published, on average, a similar number of scientific papers.

2.3.2 Managerial control and performance

2.3.2.1 Communication and performance

Giving or exchanging information, supportive and sympathetic relationships, physical connections and access to a large network of colleagues define communication. Communication processes can occur within research groups (between the leader and group members or/and among group members), between research groups (between external networks of colleagues), and between group members and their clientele. In 1956, Pelz studied a large government organisation that conducted basic medical research. He found that the performance of researchers was higher when they had: (1) daily contact with several scientific colleagues who, on average, had been employed in settings different from their own, who stressed values different from their own, and who tended to work in scientific fields different from their own; (2) frequent contact with at least one important colleague who had similar professional values; (3) a supervisor and a major colleague, where one was in the same scientific discipline and the other in a different one, rather than both similar or both dissimilar. A decade later, Pelz and Andrews (1966) reported that the average frequency of contact (conversations, memoranda, seminars) between researchers

working in a laboratory setting and their most important colleagues correlated positively to their performance. Scientists who performed at the highest level spent considerably more time on communicating with their colleagues (optimum 6–10 hours per week). Furthermore, high performance was shown by scientists who had contacts with a large number of colleagues working in their own group as well as outside their research group (but within their 'own' organisation). The positive relationships found between colleague contact and the performance of researchers appeared even after differences in experience, in supervisory status and in who initiated the contact, were taken into account. Also, Allen (1970) found that high performers (working in R&D laboratories) communicated to a larger extent with organisation colleagues, spent more time on their discussions with colleagues, and relied on more people both within their own speciality and in other specialities. By interviewing successful scientists, Sinderman (1985) found extensive references to communication and networking. It seemed that successful researchers usually have long-term cordial relationships with peers and colleagues. Relationships with peers usually begin early in the scientist's career with discussions about methodology or design of an experiment or with enquiries about poorly understood concepts. This mutual testing process continues with, for example: office or seminar discussions with colleagues; requests for informal evaluations of research proposals and draft manuscripts; and the development of co-operative research projects (Sinderman, 1985). A large quantitative study about communication and research performance has been conducted by Visart (1979). Data were collected from 10,000 individual researchers from six countries. She found that both the between- and the within-unit communication indices explained as much as 22% of the variance in recognition, 21% in applications effectiveness, 19% in R&D effectiveness, and 31% in the number of published written products. After controlling for unit size, age and experience of staff, it was shown that the percentage of variance explained in the recognition by communication predictors tended to be lower in larger research units, units with older scientific staff and units where scientific staff had more years of R&D experience. Just the opposite was found for applications effectiveness. A few years later, in another quantitative study approach, Saxberg and Newell (1983) reported that it was the unanimous opinion among directors and members of interdisciplinary research teams that communication and interaction among team members form the most critical ingredients in keeping the group and its work on course. A wide range of reporting events was suggested ranging from the most informal team discussion to a formal seminar to which interested colleagues from elsewhere may be invited. Finally, Cole and Cole (1973) used the concepts 'awareness' and 'visibility' to analyse the process of scientific communication (exchange of information between producers and audience, consisting of other scientists working in the field).

They concluded that awareness (in this case of physicists) is not greatly influenced by the individual or contextual characteristics of physicists and that communication is relatively open. Visibility, on the other hand, is strongly influenced by individual or contextual characteristics, such as scientific achievements and the location in the research system.

2.3.2.2 Rewards and performance

According to Omta and de Leeuw (1997), the challenge of research management is to create the conditions conducive to meeting the corporate goals of scientific performance as well as the scientists' need for satisfaction and motivation. Rewards used and given by local, national and international research management can motivate employees, including scientists, to perform at higher levels, and the proper use of rewards culminates in improved performance at the organisation and group levels. Figure 2 shows the path from rewards to organisation performance. Latham and Wexley (1981) tried to identify outcomes engineers and scientists would value as rewards for doing a good job. By interviewing 242 engineers and scientists they found that the five most valued rewards were: (1) seeing one's work applied, (2) receiving a salary increase, (3) having one's work put to commercial use, (4) praise from a supervisor, and (5) a monetary bonus.

Reward practices -> Motivation -> Individual performance -> Organisational performance

Figure 2: Rewards' influence on performance according to Allen and Helms (2001)

In the literature several examples of effective reward systems for R&D researchers have been reported. In 1981, Latham and Wexley conducted an experimental study in which they assessed the effectiveness of various rewards on the performance of R&D scientists. The rank ordering in terms of impact on performance was: (1) money, (2) praise and (3) public recognition. However, they reported that the lead in performance due to the money over praise was so small as to be practically insignificant. As early as in 1960, Gustad had concluded that money seems to be a motivating factor for researchers only under limited circumstances, for example when the salary of the researcher is very low compared to others. Beyond a certain level of subsistence or comfort, salary is not rated as an important reward element (Lewis and Becker, 1979). McKeachie (1979) argued that praise, prestige, salary and promotion are important, but not so much for the material gains they provide but for their ability to grant recognition of the

researchers' special expertise, intellectual ability and value vis-à-vis their colleagues. For example, being promoted indicates that one has achieved competence in the eyes of respected peers. Furthermore, McKeachie suggested that motives important for a particular researcher at one particular time may not be important at another time (also called 'life-span development theory'). In this way the indicators salary and promotion are supposed to be of particular importance in the early years of one's career. By using a meta-analytic study design Jenskins *et al.* (1998) studied the effects of financial incentives on the performance of employees (in general, not only employees working in a R&D setting). Interestingly, they concluded that financial incentives were not related to performance quality (r = 0.08, n.s.). A positive correlation was found between financial incentives and performance quantity, in which setting consistently moderated the strength of the relationship (ranging from r = 0.23 in a laboratory setting to r = 0.46 in field experiments). Fairweather (1993) demonstrated the dominance of the research and scholarship-oriented reward structure for faculties in four-year colleges and universities. He found that, regardless of the type or mission of the institution, staff in faculties that spend more time on research and publishing were paid more than their teaching-oriented counterparts. In a study among academic medical research groups, Omta and de Leeuw (1997) found that the use of material and immaterial incentives (for example, career possibilities and planning, training facilities and recognition) together explained about 10% of the variance in performance. Stimulation of international contacts can also be seen as a form of reward. Contact with colleagues in other countries gives access to new information and is important for exchanging results and ideas. Kyvik and Marheim Larsen (1994) studied the internal contacts and performance of faculty members of the rank of assistant professor or higher. A clear correlation (r = 0.44) was found in the medical sciences between the number of international contacts (measured as international conference attendance, long-term research stays abroad, guest lectures abroad, evaluation work abroad and research collaboration with foreign scientists) and the number of international publications. Omta and de Leeuw (1997) reported that the frequency of international contacts with scientists (for instance, in international congresses and workshops) accounted for 4% of the explained variance in research performance in academic medical groups. Finally, in 1998 Badawy provided an overview of the management of human resources in R&D during the past fifty years. He reported that effective reward systems for scientists consist of several elements including: a dual ladder system, a professional award system, a career planning system, a creative climate, mentoring and effective communication.

2.3.2.3 Coordination and performance

Pelz already found in 1956 that researchers in very tightly coordinated situations were so constrained that it decreased their ability to be productive. A decade later, Pelz and Andrews examined the relationship between motivation and performance and between autonomy and performance in research units (data were obtained from 500 R&D scientists in industrial, governmental and university units) that differed in levels of coordination. They found that productive scientists working in loosely coordinated groups could be characterised as: (1) highly involved and having inner motivation, (2) had a desire for self-actualisation, (3) communicated frequently with colleagues, (4) worked closely with colleagues, (5) experienced a strong competition between groups, (6) experienced a tight competition among individuals, (7) were specialised in various areas and (8) had various R&D functions. However, they also found that a relatively high level of individual autonomy was effective mainly in the middle range of situations: those who were neither very tightly nor loosely coordinated. Pineay and Levy-Leboyer (1983) compared successful medical laboratories with less successful ones (n = 155) and found that the best laboratories were those managed by supervisors whose approach was moderately free but who did formally control their researchers' work.

2.3.3 Personal characteristics of researchers and performance

Pelz and Andrews (1966) found that *personal motivation* has a moderately positive relationship with individual performance (ratings as well as actual scientific outputs). Interestingly, they concluded that the critical element in the relation between motivation and performance is not the specific source of motivation but an underlying factor of intellectual *self-reliance* – confidence in one's own ideas. A cross-national replication was undertaken by Andrews a few years later. He reported a positive relationship between the motivation of scientists and group performance (Andrews, 1979a). Pelz and Andrews (1966) also found that *diversity* in professional activities and skills of American scientists seem to be positively related to their individual performance. More than a decade later, Andrews (1979a) showed in an international comparative study of research units that diversity, as a general phenomenon, explained roughly 10% of the variance in recognition, R&D effectiveness and number of publications. The relationship between the *autonomy* of scientists and individual performance was also studied by Pelz and Andrews (1966). They found that PhDs in research labs with moderate autonomy performed somewhat better if

their interest in breadth was weak. The reverse occurred at high levels of autonomy: PhDs who maintained a broad interest in new areas performed quite well indeed, and those who were interested in detail performed poorly. Furthermore, Pelz and Andrews (1966) argued that the *creative ability* of scientists alone is not enough to elicit good performance. They found that the environment in which scientists work might affect the likelihood of their making good use of whatever creative ability they have. Working on a project or specialising in an area for a relatively short time, being part of a team in which coordination was not too high and where one had the ability to influence important decision-makers, and having reasonably good facilities for communicating new ideas to others were three situations that seemed to enhance the pay-off from the creative ability of the scientist (Pelz and Andrews, 1966). Finally, the relationships of personal traits of scientists (including sex, race, age and class), academic origin and job histories with the performance of scientists were studied mainly in the period 1966–1979. Pelz and Andrews (1966), for example, reported a continuing increase in performance (scientific contribution, overall usefulness, number of published papers and unpublished reports) with *age* of the researcher up to the early or late forties. These measures dropped sharply in the researcher's early fifties and recovered in the later fifties[30].

Cole and Cole (1967) found that the more scientists' work is recognised and used by colleagues (also called *early recognition*), the more often they continue to be productive. In addition, they showed that when scientists are ignored, their productivity will tail off. Several studies (Crane, 1965; Reskin, 1979) showed that being trained by an *eminent sponsor* is associated with higher productivity in scientists. By studying *career contacts* in obtaining a position, Granovetter (1973) pointed out weak ties as important resources for utilising possible mobility opportunities and indispensable for the integration of individuals into communities. Strong ties, in contrast, led to overall fragmentation. Furthermore, the *prestige of departments* seemed to be important for research productivity at the middle and end of their first postdoctoral decade and for the number of citations scientists received from early first-authored work (Reskin, 1979).

2.4 Conclusion

Chapter 2 describes research groups and performance. Both the actors in the environment of research groups and employing organisations or institutes frequently evaluate research during

[30] A search for factors that might account for these differences led to a set of motivational items indicating inner motivation or self-reliance (Pelz and Andrews, 1966).

different phases in the research process. One can assess research input (also called ex-ante evaluation), research throughput and research output (also called ex-post evaluation). It can be concluded that an adequate research evaluation requires several methods, criteria and indicators. As early as in 1971, Smith and Fiedler (p. 232) stated that …'*no criterion or measure currently available is sufficiently well established to stand alone'*… In addition, research quality is a broad and complex concept with different aspects (e.g. van Leeuwen, 1996; van Raan, 1985; Vroeijensteijn, 1995) that cannot be easily applied operationally in an unambiguous way. For example, one can distinguish between cognitive and methodological quality (van Leeuwen, 1996). Cognitive quality is concerned with the importance and, often, the originality of scientific findings. Methodological quality refers primarily to the accuracy with which specific methods are applied. Another example is the difference between scientific and societal quality. The concept of the scientific quality of research is based on the view of Merton (1973), who argues that the contribution to the growth of scientific knowledge is the final objective of science. Societal quality refers to the increased pressure on researchers to demonstrate user relevance and relationships with industry or utility (e.g. van der Meulen and Rip, 2000).

Chapter two also presents empirical studies concerning group environment and research performance. It can be concluded that personal characteristics of group members, managerial control and contingencies affect research performance. To be productive, research groups must employ researchers with certain personal characteristics and researchers must also work in environments conducive to research. The empirical studies show both differences and similarities. The most important differences concern the study sample and the methodological approaches. First, a number of studies used a sample composed of excellent researchers and/or research groups, while other studies used a randomly selected sample of researchers and/or research groups. Another difference was the size of the sample, ranging from about 20 to 1300. Second, the methodological basis of the studies was different: some studies employed qualitative interviews, while other studies used a questionnaire-based collection procedure. Furthermore, the underlying theories and constructs used were different. Many personal characteristics and managerial control items identified in these empirical studies are also important in the general context of organisational behaviour and management theory. This observation was also noted by Bland and Ruffin in 1992 and just recently by Ryan (2003). There is a substantial body of management and organisation-development literature that points to the importance and influence of factors such as leadership, motivation, communication, culture and diversity in achieving performance (e.g. Martin, 1998; Moorhead and Griffin, 1998). However, since1980 these newer insights have hardly ever been applied to research management. Despite the identification of

personal and managerial factors related to research performance and the observation of the similarity to organisational variables (which are related to organisational performance), a general theoretical model explaining the relationship is still lacking. On the basis of the literature as described in section 2.3, the different factors, which influence research performance, can be grouped in three layers (see figure 3). The outermost layer shows contingency items. The second layer consists of managerial control items. The centre represents the individual level. Within this level, the personal characteristics of scientists that can contribute to research performance are placed. The variables, on both the individual and organisational levels, in determining research performance are numerous and complex. As observed by Ryan (2003), the possible interrelationships and interdependencies between them are yet to be examined. Most empirical studies that examined research performance concentrated on only one or a few research management activities, contingencies or personal scientist characteristics. Studies that have taken into account multiple research management activities in combination with a number of contingency variables are scarce. In the quantitative part of this thesis (chapter 6) an attempt has been made to clarify and provide some order for the many and complex variables presented in this chapter.

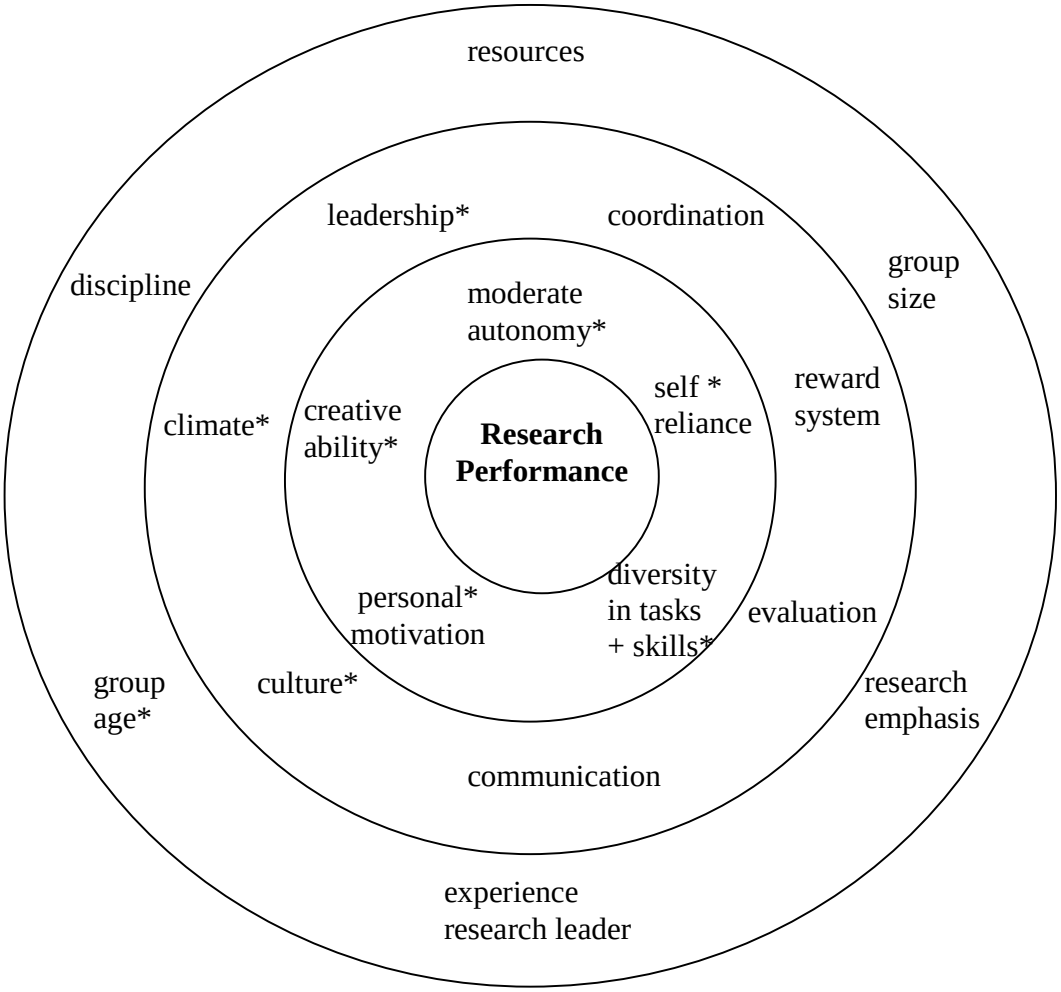

Figure 3: An overview of the variables at different levels influencing the research performance of groups. Please note that items marked with * are known in the literature as related to research performance but were not empirically investigated in this study.

CHAPTER 3: RESEARCH DESIGN

The research reported in this thesis is both qualitative and quantitative and has been developed in various stages (see figure 8). The purpose of this chapter is to elaborate on and to account for the research design of the empirical study. It starts by outlining the overall design of the project. This is followed by a discussion of the various steps in the process and by an explanation of the methodology used for each of them. In sections 3.2 and 3.3 the series of semi-structured interviews and the survey are described in more detail. The instruments of data collection and methods of data analysis are described for each of the separate steps in both the qualitative and quantitative part of the study. The results of the empirical studies are presented in Chapters 4–6.

3.1 Overall research design

The purpose of this study was to gain more insight into evaluations used to assess academic medical and health research in the Netherlands and their effects. In more detail, five research questions have been formulated (see 1.4):

1a. What different research evaluations are organised externally to assess Dutch medical and health research?
1b. What different research evaluations are organised internally to assess Dutch medical and health research?
2a. What is the impact of externally organised research evaluations on Dutch medical and health research?
2b. What is the impact of internally organised research evaluations on Dutch medical and health research?
3. Do research management activities and positive views or judgments about research management enhance the performance of Dutch para-clinical, pre-clinical and clinical research groups?

On the basis of an analysis of the literature (see Chapter 2) and a preliminary investigation[31] consisting of fifteen open interviews with experts (Dutch professors in health and management research), research administrators of pharmaceutical industries, semi-government research

[31] The preliminary investigation was conducted in 1999 in order to explore the central topics of this study.

institutes and research schools, it was possible to formulate a research model (see Figure 4). This model represents the lines of reasoning most prominently formulated or found in earlier research. The model shows that both the environment and the internal organisation of research groups steer health research. Intermediary organisations, universities, research organisations and medical faculties fund, steer, control and, finally, enhance the performance of health research groups by using research evaluations in all phases of the research process. This ranges from the granting of research resources to the evaluation of research products and outcomes. In Figure 4 the relationship between the university and the medical faculty is presented as a dotted line. This does not mean that this relationship does not exist. The research model concentrates on the possible relations that are empirically investigated in this thesis, for example international research policy is not taken into account.

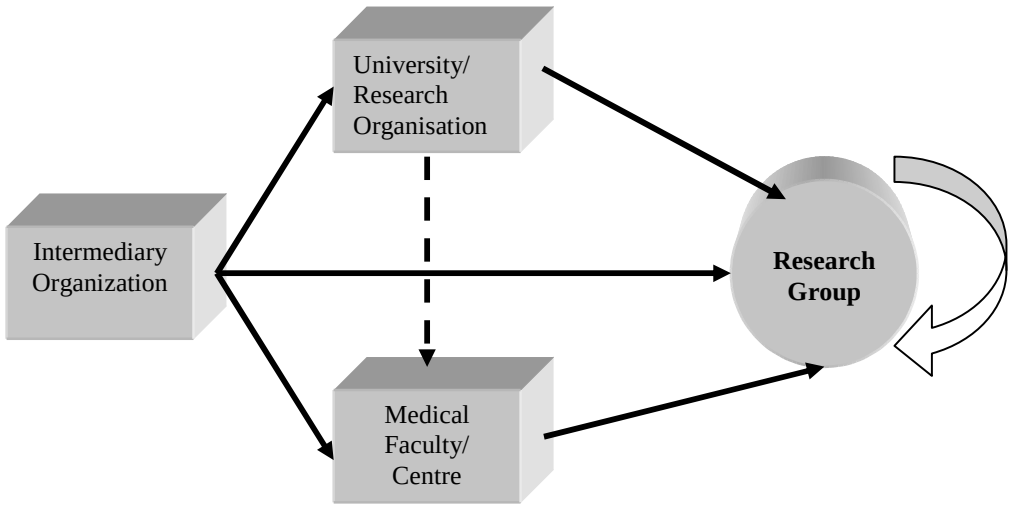

Figure 4: Research model[32]

From Chapter 2 the hypothesis can be made that the arrangement and use of management activities, such as rewards, research communication and research policy planning, by research leaders enhance the research performance of their groups (see Figure 5). Also the subjective views and judgements of research leaders about research management (for example, about such items as the importance of externally organised research evaluations) may enhance research group performance.

[32] Please note that the qualitative interviews were conducted in the period 2000-2001. Although most medical faculties were co-operating with university hospitals at that time, this arrangement was only formalised in Leiden (Leiden University Medical Center, LUMC), Utrecht (Utrecht Medical Center, UMC) and Amsterdam (Amsterdam Medical Centre, AMC). In this chapter the research evaluations that are internally organised on the level of the faculty and the medical centre (organisation level B) are only indicated, for convenience, by medical faculties.

Contingencies are factors that influence the previously mentioned elements and relationships among the elements. The size and the age of the research group are examples of contingencies. Contingencies may influence, for example, the choice of management activities and the extent to which they are implemented within research groups. Furthermore, contingencies may have a direct relationship with performance. The possible relationships among contingencies, managerial control and research performance are presented in Figure 5.

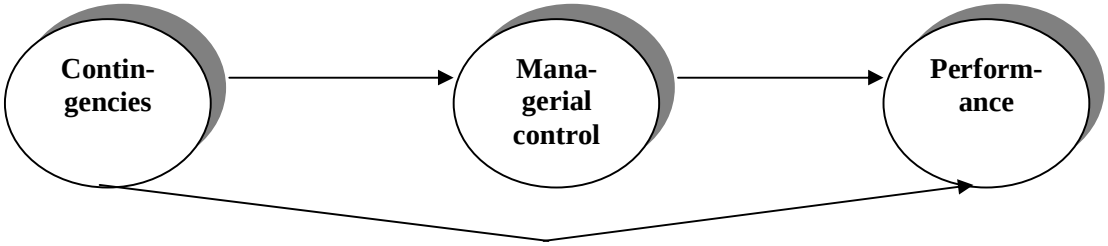

Figure 5: The internal organisation of research groups

To answer the research questions, a mixed research design was used, i.e. a combination of qualitative and quantitative techniques. The qualitative part of the study consists of three series of semi-structured interviews with successive representatives of intermediary organisations, research administrators and research leaders of academic groups who are all active in the field of health research. In total, 30 semi-structured interviews were conducted. The quantitative study consists of a survey among 412 leaders (mainly senior researchers) of academic health research groups. Also, documents were analysed during all phases of the research project. For example, organisational documents such as annual reports, strategic plans and research policy plans provide a rich source of information. They include information about mission, organisation and operation, and research management including quality control. These documents usefully complemented the interviews with background information.

Because the research was of an exploratory character, the study was organised in such a way that the first stages led to defining the research area further by finding out what is known about the topic and then developing more detailed questions about the topic for the main part of the empirical work. Two examples are given below to illustrate this further. The construction of the interview design of the series of semi-structured interviews (see Table 7), which started as early as 1999, is one example. As part of the development of the central research questions and the overall design, open interviews were held with experts (Dutch professors in health and management research), research administrators of pharmaceutical industries, semi-government research institutes and research schools. In the summer of 1999, the results of the preliminary

investigation were presented and discussed in a self-organised expert workshop in the Netherlands. Comments and suggestions from both interviewees and participants in the workshop were incorporated into the interview protocol that was eventually used in the series of semi-structured interviews. Another example is the construction of the questionnaire (see Table 7). The second part of the interview protocol used in interviews with research leaders was unstructured. The material resulting from these interviews were used to develop and to design survey questions and their possible answer categories. In addition, the survey was pre-tested by a sample of eight health research group leaders. The next sections will describe the interviews and the survey in more detail.

RESEARCH QUESTION	METHOD	EXPLANATION
1a, 1b, 2a, 2b	Preliminary investigation	15 open interviews with Dutch professors in health and management research, research administrators of pharmaceutical industries, semi-government research institutes and research schools
	Expert workshop	Presentation and discussion of results of preliminary investigation
	Semi-structured interviews	7 semi-structured interview with representatives of intermediary organisations 11 semi-structured research administrators of the eight medical faculties 12 semi-structured interviews with leaders of health research groups
3	Open interviews	12 open interviews with leaders of health research groups
	Survey (questionnaire)	Pre-tested on a sample of eight health research group leaders Sent to 412 medical research groups

Table 7: Steps in the research process

3.2 Semi-structured interviews

The main purposes of the interviews were to describe external and internal research evaluations (research question 1) that are commonly used in Dutch medical and health research and to explore their impact on research (research question 2). This section elaborates on the content of the interviews in relation to these purposes, selection of interviewees and the way in which the interviews were constructed, conducted and analysed.

3.2.1 Intermediary organisations and research administrators

In the period between February 2000 and May 2000, the first series of interviews were conducted. Seven representatives of intermediary organisations in health research, listed below, were interviewed (see appendix 3 for a detailed list of the actors interviewed):

- Advisory Council on Health Research (RGO)
- Netherlands Organisation for Health Care Research and Development Council (ZON)
- Medical Sciences-Netherlands Organization for Scientific Research (NWO-MW)
- Commission for the recognition of research schools (KNAW ECOS)
- Association of Universities in the Netherlands, Council of Medical Faculties in the Netherlands (DMW-VSNU)
- Royal Netherlands Academy of Arts and Sciences, Council for Medical Sciences (RMW)

Klasen (2000) showed that these six intermediary organisations are the most important ones for funding and steering health research in the Netherlands. These organisations regularly meet each other in both formal and informal settings to gear their activities to each other, to discuss the strategies pursued and to make work appointments. The most important starting-point of these meetings is that the missions of the various intermediaries are both respected and monitored. When needed, these intermediaries will make further adjustments together with the Council for Public Health and Health Care, the Health Care Insurance Board and the Dutch Society of University Hospitals.

Health research is conducted in the medical faculties of eight universities and in research institutes. The research administrators of the eight medical faculties were selected as interview partners[33]. Research administrators are not only involved in implementation processes of externally organised evaluations but also in the development and organisation of internal research evaluations. Research administrators of four research institutes active in health research were also interviewed. These research institutes were selected because they have clear similarities with medical faculties: not-for-profit organisations, participation in graduate health schools and strong connections with universities (for example, part-time appointments at universities). Between October 2000 and May 2001, a second series of 11 semi-structured interviews was conducted with research administrators working in medical faculties or research institutes (see Appendix D for a detailed list of interviewees).

[33] The research administrators were all willing to cooperate except for one. The internal evaluation protocol was reconsidered at the moment of investigation. Therefore, this medical centre decided not to participate in this study because the reconsideration processes had to be treated confidentially.

Both representatives of intermediary organisations and research administrators were questioned about their background, the mission of their organisation and the definition of the concept 'research quality' used by their organisation. They were also questioned about the characteristics of the evaluation procedures internally used at that time, developments that had recently been implemented in evaluation procedures and alterations to the evaluation procedures that will take place in the near future. Finally, they were asked to reflect on the impact of the research evaluations used on medical and health groups. Annex 1 gives a detailed list of interview questions.

3.2.2 Research leaders

The semi-structured interviews with research managers showed that internal research evaluations are not only executed at the level of the faculty or research institute but also at the level of the research group. Therefore, leaders of health research groups were interviewed. Between January 2001 and June 2001, twelve interviews with leaders of health research groups were conducted. The interviews consisted of two parts. The first part was semi-structured and conducted to gain insight, from the researchers' point of view, into the use of internal research evaluations and the effects of both internal and external research evaluations (research questions 1 and 2). The second part of the interview was unstructured and was used in the preparation of the development and design of survey questions and their possible answer categories (research question 3).

The interview partners were all experienced leaders of health research groups. In this thesis the definition of Andrews (Andrews, 1979b, p. 19) was used to define research groups. *'For a group of individuals to be regarded as a research unit, it had to meet three criteria: (1) The group had to have at least one recognised leader who was significantly involved in its work. (2) The group had to include a total of at least three people (including the leader) who were significantly involved in its work, and each of these people had to have been a member of the group for at least half a year. (3) The group had to have an expected life span of at least one year'.*

To select a small subset of the total population of medical and health research groups in the Netherlands that are fairly representative of the whole population, three criteria were used. First of all, the organisational setting of the research groups was taken into account. The groups interviewed were all linked to a research institute or university. In each of the medical faculties and research institutes only one research leader was interviewed. Second, the various (sub)disciplines of health research were considered. The selected research groups represented the

following disciplines in health research: genetics, (public) health research, cardiovascular disease, neurosciences and oncology. Third, the variation in research quality of health groups was taken into account to make sure that not only good groups were reached but also less successful ones. For this purpose, the results of a national external research evaluation were taken into consideration. An international assessment committee (IAC) performed this national evaluation in 1998. The IAC rated all health research on a 5-point scale, ranging from poor to excellent. The results of the evaluation were published in 1999. Annex 3 shows a detailed list of the persons interviewed.

In the first part, the interview partners were asked about their background and about the mission and size of their research group. They were also asked about the characteristics of their research evaluation procedures used at that time and about the effects that these evaluations could have on their health research group. In the second part, they were asked to reflect on a list of management activities that could contribute to a group's research quality. Annex 1 lists the interview questions and topics.

3.2.3 Data collecting and processing

The average length of the three series of semi-structured interviews was 90 min, ranging roughly from 45 to 200 min. The principal investigator interviewed all interview partners. The interview partners were invited by a letter containing a request to participate in the study, signed by both the researcher and her supervisors, and a background document to inform the interview partners about this PhD project and the aims of the interview. After a week the interview partners selected were called and interview appointments were made. A letter that contained an endorsement of the appointment and a review of the topics covered in the interviews were sent to the interview partners.

It is important to note that the choice to conduct semi-structured interviews implies that interviews differ from each other. The actual interview depends heavily on the specific expertise of the interview partner. Although, in general, all interview partners were asked the same questions (except in the interviews with research leaders, the question about management activities was added), the order of the questions and the importance of the topics differed. The interviews were not transcribed literally. A factual representation of the statements was sent to the respondents to correct possible misunderstandings. Finally, for each interview any misunderstandings were corrected and extremely confidential information was excluded. Some of

the conclusions drawn in this thesis are illustrated by interview quotes. All interviews were originally in Dutch. The quotations in this thesis were translated as literally as possible.

3.3 Survey study

The third research question I would like to answer is: 'Do research management activities and positive views or judgments about research management enhance the performance of Dutch para-clinical, pre-clinical and clinical research groups?' In order to investigate this research question a questionnaire was developed. This section first elaborates on the content of the questionnaire and then focuses on the way it was constructed, conducted and analysed.

3.3.1 Content

The survey comprised 49 questions, some of which were subdivided into a number of separate statements. It was divided into six parts (see Annex 4 for the complete list of survey questions). It concentrated successively on general questions about the respondent (part I), general questions about the research group (part II), research funding (part III), research management activities (part IV), internal research evaluations (part V) and external research evaluations (part VI).

3.3.1.1 Contingency factors

Parts I and II of the questionnaire form operational measures for the contingency factors such as 'size', time allocation' and 'age' (see Table 8).

- *'Size'* of a research group was measured as the number of scientific and technical staff working in the research group in 2002, in full-time equivalents.
- *'Work setting'* was measured by asking whether the research leaders worked in a university setting or in a non-university research institute.
- *'Time allocation'* was measured on a 6-point response format (1 = 1–10%; 2 = 11–20%; 3 = 21–30%, 4 = 31–40%; 5 = 41–50%; 6 = 51–100). Time was measured as percentage of the total working time allocated by research group leaders in 2002 to research, education, patient care and supervision.

- *'Age'* was measured in four different ways: *'research experience'* measured as the number of years of research experience of the research group leader; *'management experience'* measured as the number of years during which the research group leader had a management function; *'foreign experience'* measured as the number of years the research group leader had worked in foreign research groups; *'supervising experience'* measured as the number of years the current research group leader had been supervising the research group.

DEFINED CONCEPTS	**EMPIRICAL CONCEPTS**	**SURVEY**
Size	Staff under leadership of research leader	Question 6b
Work Setting	Working in a university or non-university research institute	Question 3
Time allocation	Research	Question 10ba
	Education	Question 10bb
	Patient care/clinical practice	Question 10bc
	Supervision of PhD students	Question 10bd
Age	Research experience	Question 2b
	Management experience	Question 2d
	Foreign experience	Question 2c
	Supervising experience	Question 4d

Table 8: Overview of the operational measures for contingencies

3.3.1.2 Managerial control

In this study managerial control of medical and health research groups is considered to be a composite of internal and external control. The focus is on internal control, which refers to control over personnel, resources and research processes. However, some elements of external control, such as communication with research group and environment, are included. In this study the subjective views and judgements of research leaders about research management (for example, about such items as the perceived importance of externally organised research evaluations) are combined with more objective measures such as the organisation of research evaluations. Operational measures for research management were developed from the fourth and fifth parts of the survey (see Table 9).

DEFINED CONCEPTS LEVEL 1	DEFINED CONCEPTS LEVEL 2
Internal control	Resource control
	Research commitment
	Time spent on internal managerial control
	Communication
	Rewards
	Research policy planning
	Internal organisation of research evaluations
	Effectiveness of pre-evaluation of research proposals
	Importance of internally organised research evaluations
External control	Time spent on external research activities
	Time spent on external managerial control
	Importance of externally organised research evaluations
	Importance of collaboration

Table 9: Defined concepts that together constitute the construct of research management

Internal control

- *'Resource control'* was measured on a 5-point response format (ranging from '1 = always short of resources' to '5 = never short of resources'). It refers to the subjective assessment of the adequacy of personnel resources (laboratory technicians, PhD students and research staff with a PhD degree) in 2002 to reach the goals and objectives of the research group.
- *'Research commitment'* was measured on a 5-point response format (ranging from '1 = disagree entirely' to '5 = agree entirely'). It refers to the subjective assessment of the research group leaders' commitment to research conducted by their research group (12 items). One example is 'I'm intensively involved in at least one research project conducted within my research group'.
- *'Time spent on internal managerial control'* was measured on a 6-point response format (1 = 1–10%; 2 = 11–20%; 3 = 21–30%, 4 = 31–40%; 5 = 41–50%; 6 = 51–100). It refers to the amount of time (measured as percentage of total working time) research group leaders allocated in 2002 to internal management activities.
- *Communication*

  'Research process communication' was measured by the frequency of research meetings within the research group. Four different research meetings were distinguished: (1) progress meetings of current research projects; (2) research presentations; (3) discussions about (conference) papers; and (4) discussions about research proposals. The scale items used a 5-point scale with anchors of 1 = never and 5 = once a week.

'*Direct communication'* was measured as a percentage of total communication. It relates to the frequency with which research group leaders used e-mail, telephone and personal talks to communicate with their research staff.

- *'Rewards'* were measured on 3-point response formats (1 = not used; 2 = used to a minor degree; 3 = used to a larger degree). This is a subjective assessment of the extent to which rewards are used to motivate staff. Four different kinds of rewards were measured:
  - '*development of research skills' (*= possibility to take national and international courses, to attend national and international conferences, to gain experience in foreign research groups, to supervise MSc and PhD students);
  - '*flexibility'* (= possibility to have flexible working hours and to work at home);
  - '*special commendations'*;
  - '*financial bonus system'*.
- *'Research policy planning'* was measured on a dichotomous response format with categories 'no' (coded 0) and 'yes' (coded 1). Research group leaders were asked to indicate if they organised discussion meetings about the research policy of their own research groups.
- *'Internal organisation of research evaluations'* was measured on a dichotomous response format with categories 'no' (coded 0) and 'yes' (coded 1). Internally organised research evaluations could take place on different levels. In the survey three levels were distinguished: research group/department, medical faculty, and university/non-university research institute. Research group leaders were asked to indicate whether *'job evaluations'*, *'pre-evaluation of research proposals'* (evaluation before proposals are submitted to external funding agencies) and '*evaluation of research output'* were internally organised on different organisational levels. Additional questions about the goals of internally organised evaluations, the frequency with which these evaluations take place, criteria used and the staff who carry out the internal evaluations were asked. These questions were analysed and are described in Chapter 5 because they form a part of the first research question.
- *'Importance of internally organised research evaluations'* was measured on a 5-point response format (ranging from 1 = worthless to 5 = very useful). This is a subjective assessment of the usefulness of evaluation results and recommendations as well as the incorporation of evaluation results and recommendations.

Table 10 gives an overview of the defined and the empirical concepts that together constitute the construct of internal control.

DEFINED CONCEPTS	EMPIRICAL CONCEPTS	SURVEY
Resource control	Personnel	Question 20
	Equipment	Question 30 (1)
	Research material	Question 30 (2)
	Data processing	Question 30 (3)
	Library service	Question 30 (4)
	Laboratory space	Question 30 (5)
Research commitment	Research commitment	Question 11
Time spent on internal managerial control	Time allocated to internal management activities	Question 10b
Communication	Research process communication	Question 26
	Direct communication	Questions 28a, 28b, 28e and 28f
Rewards	Development of research skills	Question 22a (1-4)
	Flexibility	Question 22a (6-7)
	Special commendations	Question 22a (8)
	Financial bonus system	Question 22a (5)
Research policy planning	Research policy planning	Question 29a
Internal organisation of research evaluations	Job evaluation	Question 23
	Pre-evaluation of research proposals	Question 33a
	Level of research group/department	Question 33a
	Level of medical faculty	Question 33a
	Level of university/non-university institute	Question 33a
	Evaluation of research output	Question 36
	Level of research group/department	Question 36
	Level of medical faculty	Question 36
	Level of university/non-university institute	Question 36
Effectiveness of pre-evaluation of research proposals	Effectiveness of pre-evaluation of research proposals	Question 35
Importance of internally organised research evaluations	Internally organised research evaluations	Questions 41 and 42

Table 10: Overview of the defined and the empirical concepts that together constitute the construct of internal control

External control

- *'Time spent on external research activities'* was measured on a 6-point scale (1 = none; 2 = 1–10 days; 3 = 11–20 days; 4 = 21–30 days; 5 = 31–40 days; 5 = 41–50 days; 6 = more than 50 days). It refers to the amount of time that research group leaders allocated (in 2002) to: (1) presentation of lectures, (2) attendance of conferences, (3) organisation of conferences, (4) participation in editorial boards of journals, (5) participation in audit committees, and (6) participation in assessment committees.
- *'Time spent on external managerial control'* was measured on a 6-point response format (1 = 1–10%; 2 = 11–20%; 3 = 21–30%, 4 = 31–40%; 5 = 41–50%; 6 = 51–100. It refers to the amount of time (measured as a percentage of total working time) research group leaders allocated in 2002 to external management activities.

- *'Importance of externally organised research evaluations'* was measured on a 5-point response format (ranging from 1 = worthless to 5 = very useful). This is a subjective assessment of the usefulness of evaluation results and recommendations as well as the incorporation of evaluation results and recommendations. *'DAG 1998 evaluation'* and *'research school evaluations'* were distinguished.
- *'Importance of collaboration'* is a subjective assessment. The importance of collaboration with *'international and national research groups'*, *'ministries and semi-governmental organisations'*, *'pharmaceutical industries and firms'*, *'medical clinics'* and *'family doctors'* was specified. For all items a 10-point scale was used with marks ranging from 1 to 10.

Table 11 gives an overview of defined and empirical concepts that together constitute the construct of external control.

DEFINED CONCEPTS	**EMPIRICAL CONCEPTS**	**SURVEY**
Time spent on external research activities	Time allocated to external research activities	Question 31
Time spent on external managerial control	Time allocated to external management activities	Question 10be Question 10bf
Importance of externally organised research evaluations	DAG evaluation 1998 Research school evaluations	Questions 47a , 47c Questions 47b, 47d
Importance of collaboration	Research groups Ministries and semi-government organisations Pharmaceutical industries and firms Medical clinics Family doctors	Question 32a 1-3 Question 32a 4+9 Question 32a 5+6 Question 32a 7 Question 32a 8

Table 11: Overview of the defined and empirical concepts that together constitute the construct of external control

3.3.1.3 Research performance

In the survey study six measures were used to indicate research group performance. Although there are technical and methodological problems in using an indicator for gathering and handling data, the simultaneous use of various measures gives a good description of research group performance. Two operational measures for research performance were developed from part III and part IV of the survey. The remaining performance measures were derived from public sources. The performance measures can be divided into two broad groups: input and output measures (see Table 12 for an overview).

Research output measurements

- *'Number of SCI publications'*

  A computer search was carried out to count the numbers of publications attributed to each of the (371) research group leaders on the basis of the ISI database (Web of Science). To avoid misinterpretation, the family name and the initials of the research group leader were combined with the name of the city where the research group was located. The number of papers (normal articles, letters to the editors, notes, reviews) was counted in which the research group leader was a (co-) author and which were published in international scientific journals entered in the SCI in 1999, 2000 and 2001.

  A problem with publication counts is that publication traditions vary among medical sub-disciplines, e.g. Omta (1995). Whereas in some disciplines a publication in international SCI journals is the most common way of transferring scientific knowledge, in other disciplines this is often thought to be a publication in Dutch or in non-SCI journals or books. Another problem is the difference in publication strategy between different research groups within a medical sub-discipline. Whereas in some groups the research strategy is to get articles published only in prestigious international SCI journals (journals with a high impact factor, the average number of citations in the journal, depending on the discipline), in other groups the strategy is to get as many articles published as possible.

- *'DAG score 1998'*

  The primary purpose of the Discipline Report on medical and Health Sciences Research 1998 (called DAG 1998) is to provide quality judgements of medical and health sciences in the Netherlands. Research at all Faculties of Medicine and/or Health Sciences and most non-university research institutes in this field in the Netherlands during the years 1992–1996 was evaluated. An IAC evaluated health research on the level of main research themes. An overall quality assessment was given for each main research theme, on a five-point scale, ranging from 1 (poor) to 5 (excellent). In the survey respondents were asked to write down the overall quality score of the research theme to which their group belonged.

Research input measurement

- *'External research funding'*

  To obtain comparable data about research funding that groups received from external sources, research group leaders were asked to give an estimate of the percentage of research funding (in 2002) obtained from:

  - The Netherlands Organization for Scientific Research (NWO);

- The Netherlands Organization for Health Research and Development;
- The Royal Academy of Arts and Sciences;
- Dutch ministries;
- Charity funds
- Pharmaceutical industries and firms;
- International funds.

- *'Number of research proposals submitted for NWO grants'*

  The NWO promotes scientific research at Dutch universities and research institutes and seeks to raise the quality of that research. The NWO is committed to ensuring that the level of the research carried out in the Netherlands is and remains among the highest in the world. NWO research funding is allocated through a stringent selection process based on the quality of the research proposals submitted, as well as through the provision of support to individual researchers (www.nwo.nl). In this study, the number of research proposals submitted and accepted in competition for NWO grants in the period 1999–2001 was counted for each of the respondents. A submitted research proposal was rated as valid when:

  (1) the name of the research group leader was mentioned in the research proposal as the main applicant, fellow applicant, research leader or researcher;
  (2) the research proposal was entered in Delphi, the electronic database of the NWO.

- *'Number of research proposals submitted for MW-NWO grants'*
- *'Number of MW-NWO grants received'*

  Medical Sciences[34] is one of the research departments of the NWO. The number of research proposals submitted and accepted in competition for MW-NWO grants over the period 1999–2001 was counted for every respondent. Also the number of MW-NWO grants received in that period was counted for every respondent. Both the research proposals submitted and the research grants received were rated as valid if:

  (1) the name of the research group leader was mentioned in the research proposal as 'main applicant', 'fellow applicant', 'research leader' or 'researcher';
  (2) the requested or received grant belonged to one of the following research programmes that are supported by the MW-NWO: fellowships, research supports, research programmes, clinical research training grants, PIONIER grants (a Dutch

[34] In 2001 the Health Research and Development Council (ZON) and the Medical Sciences of the Netherlands Organization for Scientific Research (MW-NWO) merged. In this study the number of proposals submitted for ZON grants and the number of ZON grants received are not taken into account.

abbreviation for Personal Impulse for Research Groups with New Ideas for Excellent Research), equipment grants, clinical research, and programmed research (e.g. nutrition and chronic diseases, memory and dementia)

(3) the grant received was entered in Delfi, the electronic database of the NWO.

- *'NWO rating'*

  The rating of the research group leaders approached was measured by means of the electronic database of the NWO, called Delfi. Rating was expressed on a 4-point scale; higher values indicate a more positive judgement (higher rating). Research leaders obtained the highest rating (score 4) when they were registered both as reviewer and receiver of grants from MW-NWO and/or NWO. Score 3 was given to research leaders who were registered as a reviewer as well as an applicant for a research proposal(s). Score 2 was given to research leaders who were registered as a reviewer but never had submitted a research proposal to the NWO. The lowest rating (score 1) was given to research leaders whose name was not recognised in the Delfi database.

DEFINED CONCEPTS	**EMPERICAL CONCEPTS**	**SURVEY**
Research output	Number of SCI publications (period 1999–2001)	Not self-reported
	DAG score1998	Question 44
Research input	Percentage of external research funding	Question 14
	Number of submitted research proposals for NWO grants	Not self-reported
	Number of submitted research proposals for MW-NWO grants (period 1999–2001)	Not self-reported
	Number of MW-NWO grants received (period 1999–2001)	Not self-reported
	NWO rating	Not self-reported

Table 12: Overview of operational measures for research performance

Most of the questions were closed. Closed questions can be answered by simple checking a box or circling the proper response from a set provided by the researcher (Fowler, 1993). An advantage of closed questions is that they facilitate the efficient comparison of respondents.
For most of the items a 5-point response format was used. A limited number of items were assessed with 2- and 3-point response formats (with the options 'yes', 'no' or 'do not know').

3.3.2 Process

The survey was conducted in the winter of 2002. As described in section 3.1, it was constructed on the basis of interviews with twelve experienced research leaders and on a literature review. The questionnaire was composed with the assistance of survey experts and was pre-tested on a sample of eight health research group leaders from five different universities and three research institutes. According to Converse and Presser (1986), pre-tests can be used to test the variation, meaning, task difficulty as well as respondents' interest and attention of specific questions. Pre-testing can also be used to improve the questionnaire as a whole, for example, testing the 'flow' and naturalness of the sections, the order of questions, skip patterns and timing (Converse and Presser, 1986). The comments and suggestions of the pre-testers, on both individual questions and the questionnaire as a whole, were used to construct the final version of the questionnaire.

The questionnaire was sent to research groups in medical faculties of the eight Dutch universities and four health research institutes. These medical faculties and health research institutes were selected because, on an organisational level, these faculties and institutes were already participating in this study (by participating in the semi-structured interviews). Names and addresses of research groups and their leaders were obtained from research managers of the medical faculties and health research institutes after the semi-structured interviews had been conducted. The lists of names and addresses were checked in the Dutch Research Database (Nederlandse Onderzoek Databank, NOD) and in PUBMED. NOD contains information on current research projects, researchers and research institutes. PUBMED is the National Library of Medicine's search service. A total of 412 medical and health research groups were approached.

The respondents answered self-administered questions. Fowler (1993) argues that the choice of a self-administered questionnaire means the choice of closed questions. With no interviewer present to probe incomplete answers for clarity and for meeting consistent question objectives, the answers to open questions will not be comparable across respondents, and they will be difficult to code. There are some questions in the survey that ask about events and behaviours that are difficult to report with accuracy. In that case, self-administered procedures have an advantage over interviews because they provide more time for thought, for checking records (Fowler, 1993) and for discussions with research group members.

The questionnaire was mailed to the research leaders of the groups (mainly professors). A mail strategy was chosen to collect the data because of the relatively low costs (time and money). Furthermore, it provides access to widely dispersed samples and samples that are difficult to reach by phone or in person (Fowler, 1993).

In order to achieve a response rate as high as possible, Dillman's tailored design method (Dillmas, 2000) was used. Dillman points out that five elements are needed for achieving a high response rate:

1. a response-friendly questionnaire;
2. up to five contacts with the questionnaire recipient;
3. inclusion of stamped return envelopes;
4. personalised correspondence;
5. a token financial incentive that is sent with the survey.

During pre-testing the eight pre-testers were asked to indicate whether the questionnaire was sufficiently response-friendly. This resulted in adaptations to the format and physical size of the questionnaire. The questionnaire was sent together with a personalised cover letter (explaining research goals and approach) and a stamped return envelope. In most cases (263, 64%) a letter of recommendation, advising the research leader to participate in this study, written and signed by research administrators of the medical faculties and research institutes, was sent together with the initial mailing. After the initial mailing, three series of follow-up mailings were sent to the research leaders. Two weeks after the initial mailing the entire sample received a reminder. A new questionnaire and a covering letter were sent to all non-respondents four weeks after the initial mailing. The remaining non-respondents received, seven weeks after the initial mailing, a personal e-mail containing a reminder and the questionnaire.

Due to budgetary restraints it was decided not to give financial incentives to the respondents.

3.3.3 Methods of data analysis

The survey data were entered in SPSS version 10. The quantitative data were subjected to different statistical analyses, such as descriptive statistics, bivariate and multivariate analysis. The bivariate procedures included *t*-tests, one-way analysis of variance (ANOVA), Kruskal-Wallis tests and Pearson correlations. The multivariate procedures included factor analysis and multiple regression analysis. The results as well as the specific operations on the data are described in Chapters 4, 5 and 6.

3.3.4 Reliability of the instruments

Cronbach's α was calculated for the individual subscales in order to find out whether they corresponded with the variables defined and to check the homogeneity, which were supposed to measure a single concept. Table 13 shows that in most of the cases (90%), Cronbach's α is sufficiently high (>0.60) to warrant confidence in the internal consistency of the scales. It should be noted that the variable 'research process communication' did not reach the threshold of α>0.60. However, every item of the scale had item-total correlations higher than 0.30, so it was decided to include this variable in the analysis of the results. In 1980, Nunnally recommended a standard of 0.70 as a satisfactory level of internal consistency.

It can be concluded that the level of Cronbach's α is sufficient to provide confidence in the reliability of the operationalisation of research management activities and subjective views and judgements about research management.

	NUMBER OF ITEMS	**CRONBACH'S α**
RESEARCH MANAGEMENT ACTIVITIES		
Research commitment	9	0.81
Reward system: development of skills	4	0.60
Reward system: flexibility	2	0.65
Adequacy of human resources	3	0.78
Research process communication	4	0.53
Direct communication	3	0.63
Time allocation of external research activities	6	0.73
Importance of DAG assessment	7	0.85
Importance of research school assessment	2	0.83
Importance of internal research assessments	2	0.79

Table 13: Consistency of the empirical concepts of management activities

CHAPTER 4: EXTERNALLY ORGANISED MEDICAL AND HEALTH RESEARCH EVALUATIONS AND IMPLICATIONS

In competition and decisions about appointments, promotion, tenure positions, allocation of research funding and publication of papers, evaluation of research is crucial throughout the scientific career of researchers and for the survival of research groups, departments, faculties and institutes. The purpose of this chapter is twofold. First, it gives an overview of externally organized evaluations that are frequently used to assess medical and health research quality in the Netherlands. Second, this chapter explores the possible effects of these externally organised evaluations on research groups[35]. The evaluation practices organised by intermediary organisations to assess Dutch medical and health research are described in section 4.1. Next, the most important recent changes in the Dutch externally organised evaluation system are described and discussed in section 4.2. The last section (4.3) describes the implications of these externally organised evaluation procedures and evaluation outcomes on research groups and research leaders. This chapter is mainly based on semi-structured interviews. Representatives of intermediary organisations, research managers in medical faculties and non-university research institutes and medical and health research leaders were interviewed. Also, official policy documents and discussion papers are accounted for in this chapter. In a few subsections statements are supported by survey data.

4.1 Intermediary organisations

Intermediary organisations are part of the research system that mediates between the interests of the government and the public on the one hand and the ongoing scientific work and its institutional arrangements on the other (e.g. Rip, 1990; van der Meulen and Rip, 1994). Examples of this are: research funding organisations, advisory bodies, standing review panels and associations of research-performing institutes. These intermediaries are also called 'buffer organisations' or 'buffers' (Fenger, 1992). The intermediary level has grown in size and

[35] Parts of this chapter have been published in Dutch (van der Weijden, Groenewegen and Klasen, 2001; 2002). See the research article 'Van beoordelingslast naar beoordelingslust? Vermindering belasting voor gezondheidsonderzoekers door externe beoordelingsystemen? [From troublesome evaluation to a pleasurable task? Reducing the burden for medical and health researchers by means of externally organised evaluation systems?] and an extensive Dutch research report 'Intermediare organisaties in het gezondheidsonderzoek' [Intermediary organisations in health research].

complexity, and in importance since 1945 (van der Meulen and Rip, 1994). Rip (1999b) argues that the occurrence of the intermediary layer of institutions between the national state and research-performing institutions is an important aspect of the modern research system. The intermediary layer is made up, to a large extent, of research funding organizations. These organisations (a composite organization or a cluster of organizations) allocate public funds for basic and strategic research which do not go to universities immediately, and which involve the research communities themselves in the allocation proces (OECD, 1992; van der Meulen and Rip, 1994). They are both *'parliaments of the scientific community'* and *'governmental bureaucracies'* (Williamson in OECD, 1992) and responsive to bottom-up and top-down pressures and policies, receiving and transmitting in both directions (van der Meulen and Rip, 1994). During the last few decades, the environment of the research funding organisations has become more complex. In the literature scholars speak of the *'transition of the research system'* (Cozzens *et al*., 1990). As already touched upon in Chapter 1, research funding organisations must relate to the changing character of the production of knowledge (see 1.3). Furthermore, public and governments become more directive, and exert pressure to receive value for money. Research funding organisations are no longer alone in the intermediary level between the state and the researchers. Not only do new strategic programmes for science and technology (with increasingly frequent cross-links with research funding organisaitons) operate at the intermediary level, trying to attract good research. Universities, also, apart from influencing their own researchers, have become more active as institutions and have moved into the market of strategic research (Rip, 1999b). Van der Meulen and Rip (1994) mentioned the ongoing trends towards co-funding arrangements as an example of complexity and transition. In addition, international funding possibilities are gradually becoming more important for (national) scientific research. For example, the involvement in EU research programmes has been the most significant trend of internationalisation in research during the second half of the 1990s (Husso, Karjaleinen and Parkkari, 2000). In the Netherlands, the Innovation Platform is the most recent actor operating at the intermediary level of the research system. It all started with the European summit in Lisbon in 2000 where the European leaders defined a new mission for Europe: to become the most competitive and dynamic knowledge economy in the world. The mission of the Innovation Platform is to strengthen the innovative capacity of the Dutch economy and to turn the Netherlands into a leading country in the European knowledge economy by 2010. The Innovation Platform was installed by Royal Decree for a three-and-a-half year period, from 1 January 2004 to 1 July 2007. By means of the platform the Prime Minister works together with the Minister for Economic Affairs, the Minister for Education and 15 leaders from business, academia and society. Together they build bridges between the

different actors involved in improving the innovation climate of the Netherlands, and help them to take action (for more detailed information see http://www.innovatieplatform.nl).

In their decision making about the allocation of and accountability for money to researchers and research groups for example, intermediaries organise evaluations. Their final decisions are based on the results of these evaluations. Several intermediary organisations are operating in the Dutch medical and health research field. In this chapter the features of evaluation procedures (goals, methods, criteria) organised by five intermediary organisations are described:

- Governmental research advice and recommendations given by the Advisory Council on Health Research (4.1.1)
- Evaluation of research proposals organised by the Netherlands Organization for Health Research and Development (4.1.2)
- National evaluation of research output organised by the Council for Medical Sciences of the Royal Netherlands Academy of Arts and Sciences (4.1.3) and the Disciplinary Board of Medical Sciences of the Association of Universities (4.1.4)
- Evaluation of research schools organised by the Research School Accreditation Committee (4.1.5)

Prior to the description of each evaluation procedure, the goals and roles of intermediary organisation(s) involved in the evaluation procedure are presented.

4.1.1 Advisory Council on Health Research

The Advisory Council on Health Research, established in 1987 by law, is a sector council and a part of the so-called cooperating sector councils (COS). The RGO advises the government, especially the Ministry of Education, Culture and Science, the Ministry of Economic Affairs, and the Ministry of Health, Welfare and Sports, on setting priorities for health research and technology development in the health sector[36]. Also recommendations are given in infrastructural matters. The task is to gear health research to social needs by means of a close interaction between government, scientists and end-users of health research. To prepare its advice, the RGO appoints committees in which members of the scientific community (the suppliers) and the end-

[36] For more information see http://www.rgo.nl/en/

users of health research (patients and industry) are represented. Usually members of the RGO[37] as well as non-member experts (researchers and representatives of both patient organisations and National Health Services) participate in these committees. The advice of the RGO, presented in report, is formulated after an exploration of the field of interest. The analysis of a social health problem, the needs and the scientific research requirements and possibilities are listed after studying the specific research by reading relevant literature and interviewing researchers. In addition, the RGO frequently organises symposia to keep in touch with researchers and physicians and to identify issues that are important for both parties. As part of the exploration of the field, the committee also studies the outcomes of external input and output research evaluations based on peer-reviewed papers and/or bibliometric analyses reports). The results of such an exploration are discussed in the committee, after which an advisory report is formulated. The recommendations are discussed by the board of the RGO and presented to the government. The government decides whether and how recommendations are adopted. In general, the advisory reports result in the development of specific research programmes or infrastructural support[38], which should be commissioned by other intermediary organisations (for example, the Netherlands Organization for Health Research and Development). One example, mentioned by one of the respondents, is research in the area of chronic diseases: *'In the beginning of the 1990s, RGO investigated the area of chronic diseases, such as rheumatoid arthritis, chronic non-specific lung disease and diabetes, and recommended the stimulation of scientific research of chronic diseases. This resulted in liquid capital of 40 million Dutch guilders'*. Another example are the *'Brede analyse gezondheidsonderzoek* (BAGO) advices in 1994 that have had great impact on the steering and funding of Dutch medical and health research. The establishment of ZonMw, in order to improve and strengthen the relationship between research and clinical practice, is a direct implication of the BAGO advices.

The work load for individual medical and health researchers is restricted. Occasionally when an advisory report is being prepared in their research area, researchers may be asked to deliver data (to respond to a questionnaire or to participate in a symposium) and/or to participate in the advisory committee. The RGO has influence in the field structure and conditions, and sometimes also on the distribution between specialties. Of the various intermediaries studied, it is the most distant organisation from the perspective of research groups and scientists.

[37] The RGO has 15 advisory members with a wide range of expertise, including representatives of research organisations, ministries and intermediary organisations (Raad voor gezondheidsonderzoek, unknown).

[38] For an overview of the advisory reports and their impact on the medical research field, see (Raad voor gezondheidsonderzoek, 2000a; 2000b)

4.1.2 The Netherlands Organization for Health Research and Development

The Netherlands Organization for Health Research and Development [39] is the national funding body responsible for promotion and innovation of health research, ranging from basic, strategic and applied research to health care (the 'knowledge continuum'). ZonMw also supports the transfer and implementation of knowledge, ensuring that knowledge is exchanged between all relevant stakeholders (researchers, professionals, patients/consumers and the general public). ZonMw is responsible for the allocation of resources to health research and development coming from the Ministry of Health, Welfare and Sports and from the Netherlands Organization for Scientific Research.

Resources are distributed in two categories of national programmes[40]. Science-driven research is stimulated by *open research programmes*, which are open to applications involving health research irrespective of medical discipline. The approach is mainly tailored to talented individual researchers or high-ranking research groups. The main indicator in the evaluation procedure is scientific quality of both the applicant(s) (individual researcher or research group) and the proposal. *Restricted programmes* provide financial resources for research on specific issues, where the interaction between research, policy and/or practice is needed to address health and health care questions in order to improve and provide innovation in health care practices (http://www.zonmw.nl). This approach is oriented towards trans-disciplinary collaboration between scientists and practitioners. Indicators used in the evaluation procedure are scientific quality and the relevance of research proposals (their merit in relation to defined targets of the particular programme). Both indicators are of equal importance. The programmes use the responsive mode, the managed mode, or a mix of both (http://www.zonmw.nl). In the responsive mode, researchers themselves define topics, which, in their opinion, are relevant to the general objectives of the programme. By publishing a call for proposals (for example in Mediator, the magazine of ZonMw and Federation of Biomedical Scientific Societies -FMWV-). ZonMw invites researchers and organisations to submit proposals. In the managed mode, researchers regarded as the most capable of achieving the programme objectives are approached, for example by sending out a restricted call for proposals to a selected group or giving a direct research assignment.

An evaluation of the brief outline of the research proposal (also called pre-evaluation) is often the first step in both the open and the restricted programme (see box 4). The pre-evaluation of the

[39] ZonMw was established in 2001 from a merger between ZorgOnderzoek Nederland and the Medical Sciences of the Netherlands Organization for Scientific Research.

[40] For more information see http://www.zonmw.nl

outline of proposals as well as the final evaluation of detailed proposals are done by peer review (individuals and panels being members of the division board). Individual peers write down their comments and rate specific aspects (criteria) on a scale. Peers evaluate anonymously, with their names being only available to members of the division board or programme committee. The applicants have the opportunity to react to the judgements of reviewers. The peer evaluations and the reply of the applicant result in priority-setting and the selection of proposals that should receive grants. ZonMw-funded research projects are further evaluated twice: halfway into (mid-term evaluation) and at the end of the funding period (final evaluation). Box 4 presents more detailed information about the evaluation objects, criteria, peers and evaluation outcomes in different phases of research projects. In addition, ZonMw (both open and restricted) programmes are evaluated. A mid-term evaluation, which is organised by the programme secretary, is focused on the processes within a programme. The scientific results of the programme are assessed within a final evaluation of the entire programme. International evaluation committees sometimes organise these final evaluations.

4.1.3 Association of Universities

The Association of Universities promotes the interests of the fourteen Dutch universities to political, governmental and community organisations. The principal goal is to strengthen the position of university education and research in society. Furthermore, the VSNU is an employers' association and develops service activities for universities (http://www.vsnu.nl). Since 1993, VSNU has been organising external quality assurance[41]. Review committees set up by VSNU screen all study programmes at all Dutch universities, make a public report of their findings and submit recommendations. With regard to evaluation of medical and health research[42], the Disciplinary Board of the Medical Sciences of VSNU has been cooperating closely, particularly

[41] The activities of the Quality Assurance department of the VSNU were taken over by Quality Assurance Netherlands Universities (QANU) in 2004. QANU works independently of universities, within the statutory framework set up for the assessment, accreditation and funding of university education and research in the Netherlands.

[42] Since autumn 2004, the Netherlands Federation of University Medical Centers (NFU) has been responsible for the mutual collaboration and policy tuning among the eight medical centres. In that respect a bibliometric analysis of Dutch medical and health research was made in 2004 by the Centre for Science and Technology Studies (CWTS). Results have also been compared ('benchmark') with international standards for medical and health research. In addition, an international committee evaluated research management of the medical centres in 2005.

in the evaluation of 1998, with the KNAW Council for medical sciences (see 4.1.4). Recently, DMW-VSNU became a member of Netherlands Federation of University Medical Centers.

Box 4: Phases in the ZonMw evaluation procedure of research projects[43]

	OPEN PROGRAMME	RESTRICTED PROGRAMME
PHASE 1 Pre-evaluation	Evaluation object: applicant + research proposal Criteria: originality, innovation and scientific relevance Peers: members of the division board Evaluation outcome: yes or no	Evaluation object: research proposal Criteria: relevance of the proposal for the programme Peers: members of the programme committee or the study group Evaluation outcome: yes or no
PHASE II Evaluation	Evaluation object: applicant Criteria: education/training, work experience, research management, grants and awards received, number of publications Peers: at least three external reviewers Evaluation outcome: 1 (poor) to 5 (excellent) Evaluation object: research proposal Criteria: scientific quality by indicators: research question, originality, approach, quality research group and feasibility Peers: at least three external reviewers Evaluation outcome: I (poor) to V (excellent)	Evaluation object: research proposal Criteria: scientific quality by parameters: research question, originality, approach, quality research group and feasibility Peers: at least three external reviewers Evaluation outcome: I (poor) to V (excellent) Evaluation: research proposal Criteria: relevance by parameters: importance of proposal for the programme, innovative character, contribution to meeting the central goals of the programme, value-for-money, extent of anticipation of knowledge transfer, and implementation (possibility of and focus on application of results in health care) Peers: at least three external reviewers Evaluation outcome: very relevant, relevant or not relevant
PHASE III Priority setting	Evaluation object: research proposal Peers: programme committee Criteria: emphasis on outcomes scientific quality research proposal Evaluation outcome: grant or reject	Evaluation object: research proposal Peers: programme committee Criteria: emphasis on outcomes, relevance research proposal. Scientific quality has also to be sufficient Evaluation outcome: grant or reject
PHASE IV Mid-term evaluation	Evaluation object: completed research Peers: programme committee and programme secretary Criteria: research activities (e.g. presentations and visits), publications and other research products (e.g. dissertations, handbooks, protocols, patents, website, folders, films, videos and CD-ROMs), future plans, collaboration with intermediary target groups or end-users, knowledge transfer and implementation, and career development of funded researchers Evaluation outcome: go/no-go decision	Evaluation object: completed research Peers: programme committee and programme secretary Criteria: research activities (e.g. presentations and visits), publications and other research products (e.g. dissertations, handbooks, protocol, patents, website, folders, films, videos and CD-ROMs), future plans, collaboration with intermediary target groups or end-users, knowledge transfer and implementation, and career development of funded researchers Evaluation outcome: go/no-go decision
PHASE V Final evaluation	Evaluation object:completed research Peers: programme committee and programme secretary Criteria: obtained research goals, research activities (e.g. presentations and visits), publications and other research products (e.g. dissertations, handbooks, protocols, patents, website, folders, films, videos and CD-ROMs, collaboration with intermediary target groups or end-users, knowledge transfer and implementation, future research activities and career development of funded researchers Evaluation outcome: approval of final evaluation report	Evaluation object: completed research Peers: programme committee and programme secretary Criteria: obtained research goals, research activities (e.g. presentations and visits), publications and other research products (e.g. dissertations, handbooks, protocols, patents, website, folders, films, videos and CD-ROMs), collaboration with intermediary target groups or end-users, knowledge transfer and implementation, future research activities and career development of funded researchers Evaluation outcome: approval of final evaluation report

[43] This box was formulated according to a description of ZON, NWO-MW and ZonMw evaluation procedures published in NWO-MW (1998); ZON (1999); NWO (2001) and ZonMw (2002). The midterm and final evaluation criteria were published in May 2004 on the ZonMw intranet.

4.1.4 Council for Medical Sciences of the Royal Netherlands Academy of Arts and Sciences

In 2001, the Royal Netherlands Academy of Arts and Sciences appointed the Council for Medical Sciences (RMW)[44], to give advice on a solicited or unsolicited basis from a scientific perspective on medical sciences[45] and to promote medical research and organisation. For these purposes, the RMW has organised, in cooperation with the Disciplinary Board of the Medical Sciences, three national evaluations of medical and health research (1988, 1992 and 1998).

The most complete evaluation took place in 1998 (KNAW, 1999) and consisted of two parts. Output of research conducted at all faculties of Medicine and/or Health Sciences and most non-university research institutes in the Netherlands in the period 1992–1996 was evaluated by peer review. First, a total of 107 research themes and 556 research programmes were evaluated by an International Assessment Committee. Experts from outside the Netherlands were selected by the IAC to pre-evaluate research within each of the main research themes. These pre-evaluations, compiled by the Dutch Committee for the Disciplinary Report (cDAG), served as input for the IAC members who performed the final evaluation. Research was finally evaluated according to four criteria: scientific quality, scientific productivity, scientific relevance and long-term viability (see Table 14 which also presents the parameters). In addition, an overall quality assessment was given for each main research theme. Second, health research was evaluated at the level of sub-disciplines by cDAG, which was composed of Dutch researchers. Each sub-discipline was evaluated by two external peers on the basis of scientific quality, scientific relevance and future aspects, thereby taking into account the publication tradition of the individual sub-disciplines and the assessment of the IAC at the level of the main research themes. The outcomes of both evaluations were expressed on a 5-point scale, ranging from 1 (poor) to 5 (excellent). In addition, strong and weak points in research were indicated and recommendations were provided for maintaining or improving research. Finally, the IAC evaluated in 1998 for the first time the research management of participating faculties and non-university institutes.

[44] The Council of Medical Sciences was appointed in 2001 as a continuation of the Medical Committee which had the same tasks.
[45] including dentistry, veterinary science and pharmacy

CRITERIA	PARAMETERS
Scientific quality	– What is the quality of the scientific output of the main research theme? – What is the international position of the main research theme? – Scientific publication in refereed journals – PhD theses – Professional results – Originality and coherence of research projects – International recognition of the members of the research group
Scientific productivity	– What is the extent of the scientific output in relation to the input in human and material resources?
Scientific relevance	– What is the significance of the research for the development of the scientific field?
Long-term viability	– What is the long-term viability of selected problem areas and the approaches followed? – What are the potential prospects of the research theme, bearing in mind national and international competition?

Table 14: Criteria used in the IAC assessment of research themes. Source: Discipline report on medical and health sciences research in the Netherlands 1998 (KNAW, 1999)

The workload for individual researchers is substantial. Participants need to submit (written) information about their research such as research objectives, research results such as a list of publications, input of scientific and support staff (in fte), a comparative analysis of output between programmes partly based on impact factors, indicators of esteem and future plans.

4.1.5 Research School Accreditation Committee

A Research School Accreditation Committee, established at the request of the government in 1991[46], is tied as an independent committee to the KNAW. ECOS is responsible for the accreditation of research schools. A research school has to concentrate on *'top-level research in the area of knowledge, often through inter-university co-operation, and needs to create a sizeable centre of excellence'* (interview with a representative of ECOS). In addition, it has to focus on post-graduate education for the next generation of researchers that should be passed on by the leading researchers in the national sub-field. In 2000, 20–25 research schools in the medical field received an accreditation. An application for accreditation of a research school has to be submitted by the Board of Governors of the responsible university. Assessment rounds are organised on an annual basis. An application subcommittee, consisting of Dutch and foreign

[46] For a historical overview of the rise of Dutch research schools, see Chapter 6 of Bartelse (1999)

researchers (mainly members of KNAW), evaluates[47] the applications and advises ECOS about accreditation. ECOS decides autonomously on accreditation for a period of six years. The evaluation is based on the performance of (1) research training, (2) scientific mission, (3) independent organisation, (4) collaboration, (5) size, (6) selection of PhDs and approval of projects, (7) PhD training, (8) post-doc policy, (9) involvement in first-phase education, (10) annual evaluation of output and accountability (KNAW, 1995; 2002). After the (first) accreditation period of six years, research schools are allowed to submit an application for follow-up accreditation. In the evaluation of a follow-up accreditation, in addition to the criteria described above, the performance of research output is also important. The research school must present the results of an external evaluation of their output by an independent committee of experts (KNAW, 2001). Within the accreditation procedure as well as the follow-up accreditation procedure specific criteria are emphasised. These criteria could differ among disciplines. In medical sciences, for example, the emphasis is on scientific quality of research groups. Nevertheless, researchers are of the opinion that the performance of the education programme is more important in the follow-up accreditation procedure than scientific research performance (including cooperation).

According to one of the respondents there is a logical explanation for this feeling: …'*it is difficult to measure the quality of researchers on the level of the research school. Moreover, this is evaluated within the first accreditation. The results of education or training are much easier to measure: it is not difficult to retrieve the number of PhD theses and, as part of the follow-up accreditation procedure, PhD students are interviewed about their experiences with the training programme. Furthermore, it is easier to influence education than research. After all, when you try to influence research you always have to deal with the difficult and autonomous researcher'*…

The development and evaluation of research schools received much attention in the 1990s. Nowadays, almost all medical and health research conducted in the Netherlands is embedded within a research school. This development does not create distinctions anymore. Among researchers doubt is raised about the quality mark of ECOS recognition and the derivation of prestige from participation in a research school.

[47] A subcommittee can call in the help of external peers in the evaluation process of application for accreditation of a research school.

4.1.6 Other external research-oriented funding agencies

medical and health researchers are also able to obtain research funding from other intermediaries which are not connected to the government (so-called private research funding agencies). Examples of this are charity organisations, special funds and pharmaceutical industries. The extent to which and the form in which these funding agencies use evaluations in their awarding process of grants differs. In general, the evaluation practices of the research funding agencies that have been awarded the official quality mark of the Dutch Central Bureau for Fundraising (CBF)[48] are comparable with the ZonMw evaluation practice as described in 4.1.2. The use of peers and the establishment of a scientific committee are some examples of similarity (Cheung, 2003).

4.2 In search of a new national research evaluation system

4.2.1 Tensions in externally organised research evaluations

In the previous section, some of the externally organised evaluation procedures[49] pose a substantial administrative burden for medical and health researchers. In practice, it seems that the various research evaluations are not in tune with one another. For example, researchers receive multiple requests to present and provide information about their research in many different ways, e.g. Obbink (2000). This increases the workload for researchers. In addition, certain externally organised evaluation practices raise some questions about the reliability of outcomes (van der Weijden, Groenewegen and Klasen, 2002). In 1996, a tripartite study group under the chairmanship of Prof. van de Kaa pointed out these problems. This study group was established under the authority of the three main Dutch organisations responsible for publicly funded research: KNAW, VSNU and NWO. Because of the high quality of research in general it was suggested that the frequency and depth of externally organised research evaluations be reduced and that, for instance, planning, requests for information and implementation of outcomes should

[48] This official quality mark is given only to charity organisations that raise and spend funds in a responsible way. One of the key conditions is that at least 75% of the money raised goes to the charity's beneficiaries. For more information, visit the CBF website (www.cbf-keur.nl) or that of the Dutch Association of Fundraising Organizations, VFI (www.goededoelen.nl). Note that both these sites are in Dutch only.

[49] Note that the interviews with representatives of intermediaries and the study of documents were mainly conducted in the period 2000-2001. Therefore, section 4.1 presents the features of evaluation of external research for evaluations that were used around this period of time.

be geared to one another (VSNU, 1996). In spite of these recommendations, externally organised research evaluation procedures did not change at all until recently. In 1999, the working group 'Quality Assurance of Scientific Research' (KWO) was established by the boards of KNAW, NWO and VSNU to develop a new national evaluation system for publicly funded research in the Netherlands, which will serve all regular public evaluation goals. In 2000, the working group published the report 'Quality Obliges', in which the new evaluation system is outlined (Werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek, 2000). In 2001, KNAW, NWO and VSNU defined their position on the proposed system[50]. This was elaborated further and resulted in 2003 in the publication of the standard evaluation protocol developed for practical use in all forthcoming research evaluations conducted under their auspices (VSNU, NWO and KNAW, 2003). In short, the evaluation protocol comprises both an external and an internal research evaluation at regular intervals. At present, this evaluation protocol has been introduced in just a few institutes[51]. The features of this new national research evaluation procedure are presented below.

4.2.2 A new national research evaluation system: self-evaluation and external evaluation

The new system aims at operating with the lowest possible burden for the researchers in such a way that researchers can spend most of their time conducting research. The main objectives of the new evaluation system are to improve the quality of research and research management (including leadership) and to contribute to higher levels of research organisations and funding agencies, government and society at large. The new evaluation system links external quality assurance to internal quality assurance. This implies that research groups initiate self-evaluations, which should give research institutes an incentive to improve the internal steering of research. The research units will produce self-evaluation once every three years. The self-evaluation alternates with preparations for the external evaluation and serves as an internal mid-term evaluation (VSNU, KNAW and NWO, 2003). It serves as a starting point for external evaluation by an evaluation committee. Research institutes and research groups evaluate their research by

[50] The report 'Quality Obliges' and the viewpoints of the KNAW, NWO and VSNU are only available in Dutch.
[51] An institute is defined in the standard evaluation protocol for public research organisations as *'a group of researchers with an articulate shared mission operating under the same management'* (VSNU, NOW and KNAW, 2003).

means of a SWOT analysis. It comprises an analysis of the internal strengths and weaknesses of the research institute/group as well as analysis of the external opportunities and threats that affect the research institute/group (see Table 15 for more detailed information). From the SWOT analysis conclusions are drawn as to the necessity of change in research objectives and strategy. In accordance with the results and conclusions of the SWOT analysis, medium and long-term goals can be set and strategy adjusted.

CRITERIA	**EXPLANATION**
Strengths	Recapitulation of the strongest aspects that emerge from documentation
Weaknesses	Recapitulation of the weakest aspects that emerge from documentation
Opportunities	Analysis of developments in science and in society at large that may effect the institute's or group's research in a positive way
Threats	Analysis of developments in science and in society at large that may effect the institute's or group's research in a negative way

Table 15: Criteria used in SWOT analysis. Source: Standard Evaluation Protocol 2003–2009 for Public Research Organisations (VSNU, KNAW and NWO, 2003)

Once every six years all publicly funded research is evaluated externally. Four criteria – quality, productivity, relevance, and vitality and feasibility (see Table 16 for more detailed information) – are used to evaluate both past performance and future plans of research institutes and research programmes. Because of publication traditions and contextual relations, elaboration of the main criteria may differ across research fields. Furthermore, the use of the main criteria should always be reviewed in relation to the mission of the research institute or group. An evaluation committee visits the institute and talks to the director, research leaders, the advisory committee and researchers of the institute asking to be heard by the committee. The committee presents its judgements on a 5-point scale, ranging from excellent to unsatisfactory and writes a public evaluation report in which the outcomes of the evaluation are outlined. A draft report is sent to the board in order that they can check completeness and consistency. In addition, the committee can organise discussions about future plans with the scientific leaders of the institute during the site visit. Recommendations about personnel policy and sensitive decisions are not part of the public evaluation report but should be drawn up in a management letter. After discussing the final report with the advisory committee of the institute, the board will draw conclusions for the future of the institute. The outcomes of the evaluation are intended to help the research organisation, the management of the research units and individual researchers to arrive at better decisions about future research, research management and policy (VSNU, KNAW and NWO, 2003).

CRITERIA	PARAMETER(S)	METHODS
Quality	International recognition Innovative potential	Qualitative: peer review by experts, including discussions with group leaders and members
Productivity	Scientific output	Quantitative: bibliometrics, technometrics, sociometrics
Relevance	Scientific impact Societal impact	Both qualitative and quantitative methods
Vitality and feasibility	Flexibility Management Leadership	Qualitative: peer review by the evaluation committee, including discussions with the institute board

Table 16: Main criteria used in the external evaluation of public research organisations. Source: Standard Evaluation Protocol 2003–2009 for Public Research Organisations (VSNU, KNAW and NWO, 2003)

Most of the researchers who responded in the survey study thought that the new national evaluation system for publicly funded research, which will serve all regular public evaluation goals and consists of a self-assessment and an external evaluation, is a positive development. Only four respondents had a negative opinion. Nowadays, each medical centre has applied the new evaluation system. Furthermore, the Dutch medical and health research has also been evaluated at national level. Under the auspices of the NFU, the Centre for Science and Technology Studies made in 2004 a bibliometric analysis of all Dutch medical and health research conducted at medical centres. In addition, an international committee has evaluated in 2005 the research management of the medical centres.

4.3 Evaluation of societal quality and impact in medical and health research

In health research the ultimate objective is to improve the health outcome of each individual. According to the Council for Medical Sciences, applied health research has a dual mission, namely a scientific and a societal one: '*it is explicitly concerned not only with the acquisition of scientific knowledge as such but also with the usefulness and implementation of scientific achievements*' (Council for Medical Sciences, 2002, p. 10). Therefore, it is not enough solely to evaluate and stimulate the scientific quality of health research. Societal quality and the impact[52] of health research is also important. As mentioned in 4.1, societal indicators are already part of the ex-ante research evaluation procedures used by ZonMw and RGO. But, the externally

[52] Societal quality and societal impact are distinguished and defined in Chapter 2 (see section 2.3.4)

organised assessments of research output in the Netherlands do not (yet) deal with societal quality or the societal impact of research separately in an explicit fashion, let alone judge it by separate criteria especially developed for this purpose. However, both the national and international assessment committees that in 1998 evaluated medical and health research in Netherlands felt a need to evaluate the research on societal quality or impact. As described by one of the respondents: *'Unfortunately, there was not enough time to evaluate both the scientific and the societal quality of health research. This was a pity, especially in the evaluation of public health research. Additional problems were the shortcomings in the evaluation methods for societal quality. As a consequence, some non-university research institutes did not want to participate in the evaluation process because in their opinion the focus on scientific quality did no justice to quality of their research'.*

During the last few years, some experimental assessments were made in the Netherlands to try to develop (disciplinary) methodologies for evaluating the societal quality of research. The health research field trial based on visitations (Begeleidingscommissie Experimentele Visitaties Gezondheidsonderzoek, 1994), which explicitly included an assessment of the societal and/or applied value, is the first example. Other more recent attempts that should be mentioned here are the development of the societal quality research profile in health research (Spaapen, 1995), the development of the research embedment and performance profile, and stakeholder analysis applied to agricultural sciences (Wamelink and Spaapen, 1999) and pharmaceutical research (Dijstelbloem, Spaapen and Wamelink, 2002). In these methods the importance of the comparative feedback is emphasised; the results of the assessment of societal quality must be interpreted in relation to its research mission. Interestingly, the report 'Quality Obliges' from the KWO working group did not examine at all the possibilities for evaluating the societal quality and impact of research. The importance of societal quality and the impact of research was explicitly mentioned for the first time in the standard evaluation protocol; it was elaborated further as part of the criterion 'relevance'. Recently a working group established by KNAW developed a methodology that could measure the societal impact of health care research outcomes. Table 4 shows the criteria and indicators of the societal impact of research output developed by this KNAW working group. The KNAW working group suggests implementing the evaluation of the societal impact of applied health research within the new national research evaluation system: *'Research institutes and groups can be asked to list and describe the indicated 'non-SCI research papers' output they consider relevant in their societal mission. They can also be asked to present other data as to the potential societal impact of research results'* (Council for Medical Sciences, 2002, p. 28). The working groups emphasise that it is important that both the

scientific quality and societal impact are assessed together by a single external review committee. Also the feedback of stakeholders is important. Therefore, the working groups propose that the external review committee may consult a panel of stakeholders, such as professionals, patient organisations, health care institutions, policy makers and research clients. This proposed methodology was presented and discussed with international researchers, experts and policymakers in the field of quality assessment in an invitational meeting on 19 June 2001. In this meeting it was concluded that the new national research evaluation system offers ample possibilities for including an evaluation of the societal impact of applied research. Most criteria listed in Table 17 can be easily implemented as part of a self-assessment (Council for Medical Sciences, 2002). In the survey study medical and health researchers are asked what they think about the incorporation of societal indicators in external research evaluation procedures? Half of the responding researchers (n = 82) think that incorporation of societal indicators will stimulate researchers to improve their performance not only for scientific quality but also for societal quality. Almost one third of the responding researchers (n = 47) do not think this would be the case.

The currently emphasised developments in the measurement of societal quality are expected to acquire a firmer position in the spectrum of evaluations methods.

4.4 Implications of externally organised research evaluations

The evaluation procedures of the most frequently used externally organised research evaluations in medical and health research have already been outlined above. The question arises if these evaluations have any impact on the research groups concerned. According to the majority of the research leaders interviewed, the notion that assessment of research quality is important for the survival of research groups is growing: *'It is important to stress the distinctive features of your research and your group in order to obtain external research funding'. 'I feel more and more that the competition between researchers is increasing'.*

CRITERION	INDICATOR
Content analysis	Professional publications
	Treatment guidelines and protocols
	Policy documents
	Cochrane library
	Textbooks
	Teaching materials
	Lay publications
	ICT and software
Citation analysis	Citations in scientific publications (both SCI and non-SCI)
	Citations in professional journals, policy documents, protocols and guidelines
Authorship	(Co)-authorship of documents mentioned above under 'content analysis'
Products	Health-care technology and services
	Instruments, programmes, methods for (assessments or implementation of) care
Funding of research	(Semi) governmental funding
Publicity	Presentation for a non-scientific audience
	Fact sheets
	Public media
	Internet
Memberships	Membership of a committee issuing a policy document or a treatment guideline
	Membership of an advisory committee
Teaching	Contributions to initial and post-initial education of health-care professionals based on research output
Implementation strategy	Membership of advisory committees
	Interactions between researchers and public administration
Independence	Operationalisation of research questions
	Research methodology
	Analysis and publication of results

Table 17: Criteria and indicators of societal impact of research output. Source: The societal impact of applied health research: towards a quality assessment system (Council for Medical Sciences, 2002).

This section describes the implications of externally organised research evaluations in medical and health research. First, the implications of the ZonMw research proposals evaluations are presented as an example of the impact of ex-ante evaluations. Next, the implications of ex-post evaluations are illustrated by studying the impact of the DAG 1998 evaluation. Finally, the implications of the ECOS accreditation of research schools are presented. The implications of these externally organised research evaluations are analysed from different viewpoints. Data were mainly collected through open, unstructured interviews with representatives of intermediaries, research managers in medical faculties and non-university research institutes, and research leaders. The implications of the DAG 1998 evaluation were studied most intensively. Data were collected in both the interview and the survey study and were analysed from three different viewpoints: (1)

on the intermediary level, (2) on the faculty/non-university institute level, and (3) on the level of the research group.

Before the empirical results were presented, it should be noted that none of the respondents had conducted any research on the possible implications of procedures and outcomes of research evaluations on research groups in the medical and health field. The implications they report were based on their own observations and experiences. In addition, the implications cannot be considered separate from the context in which science is embedded.

4.4.1 Implications of ZonMw ex-ante research evaluations

The evaluation procedures of research proposals and the allocation of the second flow of research funding have large implications for health research. As early as in 1950 the precursor of NWO, ZWO, was established to stimulate – particularly fundamental – research in the Netherlands. The term 'research quality' was introduced and is now recognised and used in the whole Dutch medical and health research community. Also a few non-university research institutes, medical faculties and research schools copy parts of the external evaluation procedures and incorporate them into their own internal ex-ante research evaluation. An example of this is the allocation of the internal research budget to PhD proposals; this will be discussed in greater detail in subsection 5.1.3. According to one of the representatives of ZonMw, the use of external peers and a peer review committee are the most important elements that are copied from the NWO procedures. Also, the addition of societal criteria in the evaluation procedure of restricted programmes influences health care researchers. As one of the representatives of ZonMw had experienced: *'…On the one hand, health care researchers are becoming positively motivated and enthusiastic to pay attention to the societal quality or impact of their proposals…On the other hand, researchers who have never performed fundamental research are, to some extent, forced by the evaluation procedure to pay attention to the methodological aspects of the research proposal…'*

Not only the procedures and methods used to evaluate research proposals, but also the evaluation outcomes and recommendations are affecting medical and health research. First, it seems that the outcomes of the evaluation of research proposals organised by ZonMw in both open research programmes and restricted research programmes have implications for the internal operations of medical faculties and non-university research institutes in the distribution of internal funding. The research proposals that do not receive funding in the open ZonMw research programmes are

generally evaluated as 'very good' (score 2). Because of the low external success rates (10–15%) the applicants have an increased chance of obtaining internal research funding. Second, one of the ZonMw representatives indicated that the evaluation of research proposals results in a high rejection rate of proposals submitted by research groups from non-university research institutes or medical faculties that have hardly any internal quality assurance, if at all. According to this representative, these research groups are improving, for example by establishing an internal pre-evaluation of proposals and/or by asking for help from high-quality groups in the process of writing proposals. Finally, both representatives of ZonMw think that restricted research programmes stimulate and steer researchers to perform research in a particular research field. According to one of them, this influence is even larger than the amount of money that can be spent by ZonMw: '*Experience shows that the stimulation of a research change or a new development is, in universities, a tardy process. It is essential for ZonMw to invest money slowly in order to stimulate discussion about the research change or development in university boards. Once the new research subject is rooted in faculties the new research activity should continue even when the ZonMw stimulation would be reduced'.*

In sum, the implications of research proposal evaluation extend beyond the funding process itself. The views on research have changed.

4.4.2 Implications of ex-post DAG evaluation 1998

This subsection explores the impact of the national evaluations of medical and health research that were organised in 1998 by RMW (KNAW) and DMW (VSNU). As stated in 4.1, the disciplinary evaluations are the most extensive external ex-post research evaluation in the medical and health field. It appears that both RMW and DMW have never performed research into the implications of the various disciplinary reports, but are highly interested in their working. The representatives of intermediaries emphasise that the outcomes and recommendations of DAG are intended for internal quality assurance of medical faculties and non-university research institutes. In addition, the disciplinary report is used as a reference book by the Dutch government. Therefore, research managers of medical faculties and non-university research institutes were interviewed. These eleven research managers, with one exception, were all involved in the DAG 1998 evaluation. This makes it possible to compare their answers. In addition, the implications of the DAG 1998 evaluations are also described here (both qualitatively and quantitatively) from the researchers' point of view. The survey study shows that the majority of research group leaders

who responded (74.1%, *n* = 119) were involved in the national disciplinary evaluation of 1998. In general, these research groups seemed to be of good quality. Half of them belonged to a main research theme that was rated as 'very good' or 'excellent'. In comparison, 15% of the main research themes were rated as excellent, 46% as good, 36% as satisfactory and 3% as unsatisfactory. Table 18 gives more detailed information about the evaluation score of both the research themes covered by DAG 1998 and the survey respondents.

DAG 1998 EVALUATION SCORE OF RESEARCH THEMES	PERCENTAGE OF RESEARCH THEMES	PERCENTAGE OF RESPONDENTS
Poor (1)	0	0 (n = 0)
Unsatisfactory (2)	3	1.7 (n = 2)
Satisfactory (2–3)	36	10.1 (n = 12)
Good (3)	Score 3 to 4–5: 46	10.1 (n = 12)
Good to very good (3–4)		10.9 (n = 13)
Very good (4)		11.8 (n = 14)
Very good to excellent (4–5)		23.5 (n = 28)
Excellent (5)	15	20.2 (n = 24)
Missing data		11.8 (n = 14)

Table 18: DAG 1998 scores of the main research themes compared with the DAG 1998 scores of survey respondents.

Direct implications perceived by research managers

In general, it seems that the outcomes and recommendations of the DAG 1998 did not have very much impact on the various faculties and institutes. In only two medical faculties did the DAG 1998 results and recommendations result in some direct changes in local policy. In one internal research institute two divisions were evaluated as unsatisfactory. As a consequence, the two divisions were restructured. According to the research manager this decision was made by the internal research institute itself, not by the board of the medical faculty, because the board holds the opinion that each research division should deserve a chance of success: …*the stronger divisions must help the weaker ones. They should not be restructured on the basis of external advice'*… In the other medical faculty, the dean talked to the leaders of research programmes judged as satisfactory (score 3) or unsatisfactory (score 2). A written agreement was made for each research programme, which contains arrangements (*'mainly to reduce the number of research themes'*) to ensure that research improves. After some time, the dean will evaluate the implementation of these plans.

In two other medical faculties the results of the DAG 1998 were compared with the results of an internal research evaluation[53]. In one medical faculty, a strength-weakness analysis was made as part of the internal evaluation of research procedure. Results were expressed on a 4-point scale: strong (score 1), weak (score 2), opportunity (score 3) or threat (score 4). In the other medical faculty, the DAG 1998 outcomes of research programmes were compared with the results of an extensive internal evaluation of research programmes. Further analyses were made when the results of both evaluations conflicted; in this comparison process the results of the internal research evaluation were emphasised.

In the remaining three medical faculties and three non-university research institutes involved in the disciplinary evaluation, the DAG 1998 results did not have a direct effect on their internal research policy because of good evaluation outcomes. As some respondents argue: *'Our research is of high quality and our faculty holds the opinion that such high-quality research groups are able to steer themselves'*

However, the most important reasons given by research managers are problems experienced in the DAG evaluation procedure. Five different problems were pointed out: the evaluation units, time spent on research, selection of peers, communication and evaluation outcomes:

– Evaluation units. In the past, RMW formulated fourteen clusters. medical and health research carried out in the Netherlands has to be subdivided into these clusters. In presenting the research information over 1992–1996 to the RMW and the VSNU, these research clusters had to be used. Interview data shows that a few clusters were not recognised or used by research managers. As a consequence, some research programmes were not evaluated separately. This resulted in a distortion of the quality of some research clusters, as the following quote illustrates: *'Unfortunately, genetics, which is a strong research area in our faculty (according to internal research evaluation outcomes), was not evaluated separately because it was presented in the neurosciences cluster. Our mistake: we never use the genetics cluster'.*

– Time spent on research. Research programmes can have several goals, for example to conduct research, to educate students and to deliver patient care. The proportional amount of time spent on each of these activities differs among research programmes. According to one of the research managers, this has been insufficiently taken into account in the DAG 1998 evaluation procedure.

– Selection process of peers. The second part of the disciplinary report 1998 presents the evaluation of sub-disciplines participating in research programmes. These disciplines were evaluated by a national peer committee. Due to the national character of peers, a number of

[53] See Chapter 5 for more information about the internal research evaluation procedures of medical faculties.

research managers have doubts about the objectiveness of this evaluation: *'Roughly speaking, the sub-disciplines were evaluated by colleagues. Therefore, the outcomes should not be taken very seriously'. ' …it's an old boys' network…'. 'One research programme of our faculty traced the name of one the peers and made some comments on the evaluation outcomes'.*

– Communication. Some failures of communication between participants and organisers of the DAG 1998 evaluation have been reported. As one research managers states: *'...our faculty has set up a document about our research management and sent it to RMW. We never received the final results of the research management evaluation. We have called several times to RMW but this did not help. This is a pity because we have spent a lot of time on writing the document, and the evaluation report included some good points'.*

– Evaluation outcomes. The whole evaluation procedure took much time – almost two years – before the results were published. In the DAG 1998 assessment, medical and health research in the period 1992–1996 was evaluated. At the time of publication (winter 1999), the research outcomes seemed to be already outdated. Some respondents wonder whether the liabilities *('amount of time and energy')* are not higher than the assets *('results are often already known to the board of our faculty').*

Direct implications perceived by research leaders

The survey study shows that research leaders of medical and health groups, in their turn, hold different views on the usefulness of the results and recommendation of the DAG 1998 evaluation. 28% of the research leaders considered the outcomes and recommendations as very useful for their own research groups. About the same proportion (29%) considered the outcomes and recommendations as not useful at all. In addition to views on usefulness, views on the possibilities of implementing the DAG 1998 results and recommendations also differed. 24% of the research leaders reported (very) intensive implementation of the results and recommendations. According to 38% of the research leaders, the DAG 1998 results and recommendations had not been implemented within the research policy of the medical faculty or research institute in which their research groups were operating. Also the twelve experienced research leaders, who were interviewed qualitatively, share this opinion as illustrated by the following quotes: *'People talked more about the DAG disciplinary report compared to the actions that have been taken. For example, research groups that received a low evaluation score brought up extenuating circumstances'. 'Our faculty did too little with the results. No extra money has been invested in both 'good' groups to maintain quality and 'poor' groups to improve quality.'*

Research leaders reported various reasons to explain why, in their opinion, the DAG 1998 results have not been implemented very well. Five different problems were pointed out with regard to the evaluation units, evaluation criteria, evaluation outcomes and communication.

– Evaluation units. Almost one third of the respondents in the survey study indicated that research was evaluated on the wrong level. For example, research was not evaluated at the level of the research group.

– Evaluation criteria. The survey study shows that almost 25% of the respondents thought that the results and recommendations were not relevant due to the use of wrong evaluation criteria.

– Evaluation outcomes. A few research leaders who were interviewed qualitatively considered the results as irrelevant because the evaluation procedure took so much time that the results were already outdated when they were published. Also, 30% of the quantitative respondents shared this opinion. As one of the research leaders said: '*…evaluated research has been conducted years ago…'*

Also, a few research leaders, both those qualitatively and quantitatively approached, claimed that the results were unreliable. As one research leader further explained: *'…our institute decided on its own what parts of the research input and research output should be externally evaluated…'*

– Amount of information. Most of the research leaders who participated in the survey study shared a negative view of the amount of information they received from the RMW and the VSNU about the DAG 1998 evaluation procedure. Almost half of the research leaders received little or no information at all about the procedures. One third of the research leaders indicated that they were poorly informed about the evaluation procedure. Only a small number of the research leaders thought they have been informed adequately. The hypothesis can be made, although this was not studied, that the shortage of information has reduced the involvement of researchers and has contributed to some extent to the relatively low impact of the DAG 1998 results.

Indirect implications perceived by research managers and research leaders

Although most researcher leaders and research managers did not experience many direct implications in the internal research policy and research agenda (see Table 18 for an overview), a few indicated that the DAG 1998 disciplinary report had indirect implications. For example, the outcomes and recommendations served a political goal, as one research manager emphasised: *'…they are used by the board to legitimize an internally announced action…'*

Also, the outcomes are important for the public at large to stress the distinctive features of the research programmes or research groups. As one of the research managers argued: *'…they may attract people to apply for a job at our department…'*

IMPLICATIONS	EXAMPLES
Direct changes in local (faculty) policy	Reorganisation of units Agreements to improve research
Comparison of outcomes of internally organised research evaluations	Strength-weakness analysis
Indirect implications	Legitimisation of internally announced action(s) Enhancing the attractiveness of the medical faculty and/or research group

Table 18: Implications of the ex-post DAG evaluation 1998

4.4.3 Implications of research schools

Qualitative interviews with both the representative of ECOS and research leaders showed that accreditation of research schools has implications for Dutch medical and health research. First, the EOCS accreditation procedure demands that decisions about research and education made in the past and to be made in the future are considered carefully. Research schools are also obliged to consider their past and future decisions about spending money. In addition, the follow-up accreditation procedure requires that self-evaluation is accomplished. According to the representative of ECOS, this reflection can result in a more effective way of allocating resources. Second, to submit an application for accreditation, research schools have to make some rules about both the purchase and use of research equipment. Furthermore, PhD courses should be developed. This increased coordination between research groups seems to be an important positive consequence of the accreditation procedure.

Furthermore, the accreditation itself can have implications. In the past, an accredited research school obtained extra research funding from the Dutch government. This benefit of accreditation has already faded away due to the expansive growth of the number of accredited research schools. According to the representative of ECOS, some research schools did not submit an application for follow-up accreditation because it did not yield profit.

4.5 Concluding remarks

Externally organised evaluations of medical and health research

It appears that Dutch research groups in the medical and health field are experiencing a lot of different externally organised research evaluations. These evaluations take place in different phases of the research process. The goals of the external research evaluations vary widely. External evaluations are used to allocate funds to research (ZonMw and other external research funding agencies), to mark the quality of research (ECOS), to advise about research (RGO, RMW, VSNU), to direct research (RGO, ZonMw) and to promote research (all). All the intermediaries examined, with the exception of RGO[54], identify themselves within the definition of intermediary organisations used in this study. Former studies about research systems, as described in Chapter 2, have shown that research systems in industrialised countries can be analysed in a three-level system. Figure 6[55] shows the relations found between intermediary organisations, government and research institutes in the Dutch medical and health field. ZonMw receives from ministries an annual budget to stimulate medical and health research (red arrow). Linked to this budget, ZonMw receives assignments from the ministries (blue arrow) to promote research focusing on a particular health care problem. The government can also ask the RGO for advice on priority setting for health research and technology development. The external research evaluations organised by ZonMw, ECOS, RMW and VSNU as well as the explorations of the RGO are based on research information. Information on both research input and research output is received on the research performance level (purple arrows). Advice and recommendations from the RGO are communicated to the government (green arrow). The outcomes of research proposal evaluations, DAG output evaluations and the evaluation of research schools are communicated to both the governmental and the research performance level (green arrows). In the case of the allocation of resources, ZonMw grants are provided to researchers. Since RGO advice often results in the development of research programmes, which are elaborated further by ZonMw, it can be suggested that RGO indirectly advises ZonMw (yellow arrow). Finally, identification of

[54] As one of the respondents further explains: …'the RGO is an independent organization. It is not responsible for any policy, nor does it allocate any funds or manage any research'...Consequently, in figure 7 the RGO is put on governmental level.

[55] Throughout the empirical (interview and survey) study the interpretation of the new national research evaluation system (as presented in 4.2) was intensively discussed. It was not certain how the new national research evaluation system would fit into this figure. Therefore, it was decided not to include the new national research evaluation system in figure 7.

important scientific developments is one of the traditional goals of the RMW. For that reason, ZonMw also obtains indirect advice from the RMW.

Externally organised research evaluations of medical and health research in the Netherlands are all based on peer review. In general, a peer evaluation protocol is used in the peer evaluation process, comprising a description of objectives, main criteria, planning and procedures that are used in the evaluation. The selection of the peer evaluators is probably one of the most important steps in research evaluation. A peer must not only be competent to carry the evaluation but also be completely independent of the evaluation object (e.g. researcher/research group/programme/institute) involved. In the evaluation of scientific quality of research, peers are invariably scientists. The question of societal relevance, which is becoming increasingly important in the evaluation of Dutch medical and health research, cannot be grafted onto the traditional peer review process. Because research activities are not only aimed at colleague scientists but also at other stakeholders, traditional peer review processes in science can be extended by including stakeholders other than scientists in the evaluation process of science. The Medical Committee of KNAW, for example, proposes the consultation of patient organisations, health care institutions, policy-makers and research clients in the new external ex-post evaluation (Medical Committee of the Royal Netherlands Academy of Arts and Sciences, 2002). External output evaluations of research schools and disciplinary output evaluation practices are conducted by peer committees. Also, a peer committee will be used in the new external evaluation system. Generally, a peer committee consists of a chair, (inter)national members (the 'peers') and supporting staff. A self-evaluation is used as an input of the externally organised ex-post evaluations. In general, the self-evaluation is sent together with other relevant material, such as an evaluation protocol, background information and a list of additional questions that need to be answered by the committee. In the new external evaluation system, the peer committee will visit the research group and institute for a few days and meet the research director, the research leader and researchers. Finally, externally organised ex-post evaluations are completed by publishing a (public) report on the evaluation outcomes and sometimes also a confidential management letter about (personnel) policy and sensitive decisions.

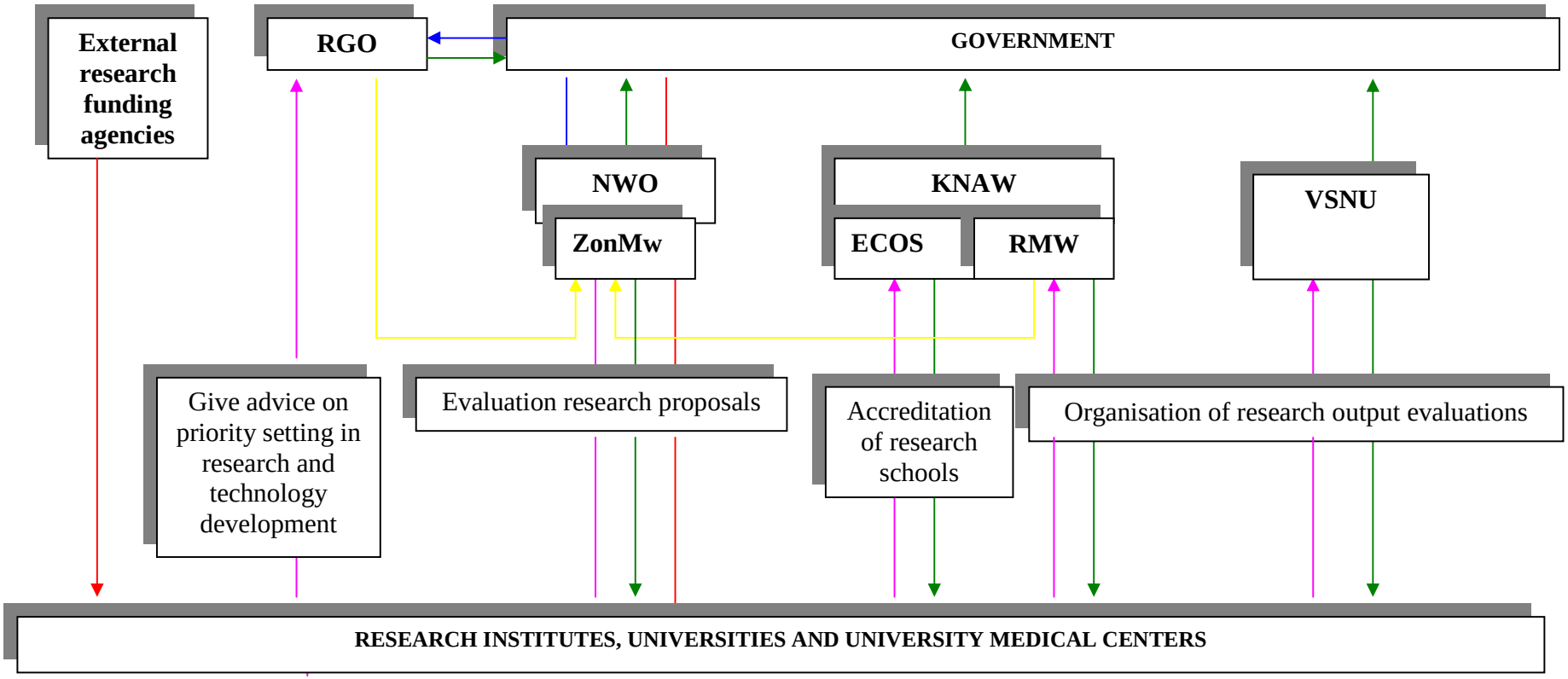

Figure 6: Relations among different actors operating in the Dutch health research system

Research funding
Information
Indirect recommendation
Assignments
Communication about research evaluation outcomes

Implications of externally organised research evaluations

In conclusion, it appears that none of the intermediary organisations conduct studies into the implications of external research evaluation. It is remarkable that, given the attention to externally organised evaluations as discussed in this chapter and knowledge accumulation in that area, intermediaries have so little information available on adjustment in internal work processes.

This study shows that both ZonMw evaluation practices and evaluation outcomes have a direct and high impact on Dutch medical and health research. Parts of the ZonMw evaluations procedures are used in internally organised ex-ante evaluations. In addition, ZonMw outcomes are highly esteemed and are used as quality marks in internal processes for the allocation of resources.

According to research managers and research leaders, the operation of the DAG 1998 evaluation was not flawless. As a consequence, the impact of the DAG 1998 evaluation on the internal research policy of medical faculties and non-university research institutes is relatively low. Both the interview and the survey study also reveal that DAG 1998 had little impact on medical and health researchers. Research managers as well as research leaders report that they experienced three main problems with the DAG evaluation procedure. First, the evaluations units are not recognised. Second, communication (frequency and intensity) between the research level and both organisers (RMW and VSNU) seems to be unsatisfactory. Third, the evaluation procedure took so much time that the evaluation outcomes and recommendations were outdated when they were published.

It can be concluded that the impact of the accreditation of research schools on medical and health research is modest. In the nineties, the number of research schools that obtained accreditation strongly increased. Thus, accreditation is no quality feature any more. Nowadays the number of ECOS research schools is decreasing. In addition, financial incentives are not given any more to research schools.

Finally, we have seen that that most of the research leaders who participated in the survey study attach higher value to the outcomes of ex-post evaluations that are internally organised. Internally organised research evaluations and their impact will be discussed in greater detail in Chapter 5 of this thesis.

CHAPTER 5: INTERNALLY ORGANISED MEDICAL AND HEALTH RESEARCH EVALUATIONS AND IMPLICATIONS

In Chapter 4 the external research evaluation practices organised by intermediary organisations are presented and discussed. It was concluded that medical and health research groups are subject to various external research evaluations which take place in different phases of the research process. As indicated in Chapter 1, research evaluations are also internally organised as part of the local research policy. Four organisational entities can be specified: research organisations (universities and non-university research institutes), medical faculties and medical centres[56] and research groups. The procedures developed and introduced by these entities to evaluate research are called internal research evaluations. The first part of this chapter (section 5.1) discusses the internal evaluations of medical and health research procedures that are developed and used. In the first phase of research (research input), it is essential to attract, employ and stimulate staff and to obtain enough research funding (see Figure 7). In that process research input evaluations are concentrated on the assessment of staff, research project proposals and research equipment acquisition. In the last phase of research, output is generated. For example, papers are published, including those in SCI journals, and/or patents are obtained. In output evaluations the research output generated is assessed. This chapter concentrates on the features of evaluation procedures (goals, methods and criteria) that are internally developed and used for assessing staff, research proposals and research output. The procedures that are internally developed and used to evaluate research throughput are not the subject of this chapter. It should be stressed at this point that it is remarkable that – given the attention to external research evaluations and knowledge accumulation in that area – so little information is available on the degree to which these evaluation procedures are translated within the internal organisation.

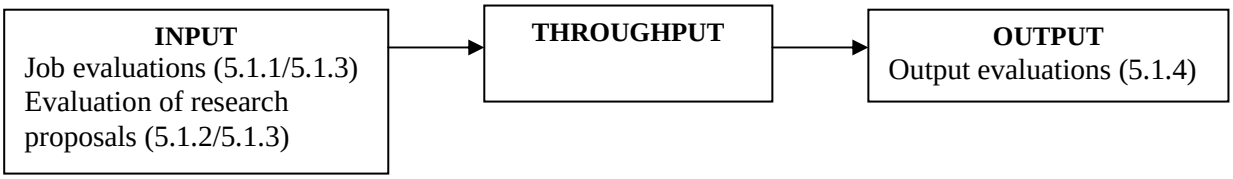

Figure 7: Evaluations are internally organized in different phases of medical research

The second part of this chapter (section 5.2) describes, from two different viewpoints (research managers and research leaders), the impact on research groups of output evaluation practices organised by medical faculties and non-university research institutes.

[56] see footnote 32

This chapter is based on both an interview study (2000-2001) and a survey study (2002). Semi-structured interviews were administered to research managers from medical faculties and centres and non-university research institutes. Also research leaders were interviewed. In addition, 160 medical research leaders participated in the survey study, viz. 127 respondents working in a university setting and 33 respondents working in a non-university research institute (see tables 40–41 in Chapter 6 for more details).

5.1 Internally organised research evaluation procedures

5.1.1 Job evaluation

The interview study as well as the survey study shows that most research groups organise official job evaluations with their staff on a regular basis[57]. Of the twelve research leaders interviewed qualitatively, only one does not organise internal job evaluations. Nine research group leaders evaluate their members yearly. Two research group leaders report an internal job evaluation once every two years. These results are supported by the outcomes of the quantitative approach: 93% (n = 149) of the responding research group leaders indicate that they organise job evaluations to assess their staff. Most research leaders (72.8%, n = 110) evaluate their employees on an annual basis (see Table 19).

FREQUENCY	**PERCENTAGE**
Every six months	4.6% (n = 7)
Once a year	72.8% (n = 110)
Every two years	20.5% (n = 31)
Every five years	0.7% (n = 1)

Table 19: Frequency of job evaluations. Three research leaders (1.9%) did not respond to this question.

As is shown in Table 20, almost all research leaders focus on the assessment and improvement of both skills and output of their staff. In order to reach individual research goals, yearly agreements between the employee and the research leader are made in 71.3% (n = 107) of the groups. The agreements as a result of the previous assessment are evaluated and adjusted wherever necessary. In many research groups employees are allowed to reflect upon the leadership style of the supervisor. In more than half of the research groups the internal job evaluations are also organised to develop and/or maintain relations between the employees and the supervisor. No significant differences were found in the

[57] It must be stressed here that this question is asked independently of the current collective labour agreement. So, it is not known whether research leaders organise job evaluation and use human resource management criteria because this is a compulsory part of the collective labour agreement, or because they want to develop an additional recruitment policy for the benefit of their research groups.

objectives of job evaluations in para-clinical, pre-clinical and clinical research groups. Finally, some research leaders and research managers qualitatively indicate that the results of job evaluations are also used in decision-making about permanent research positions (for example, about the appointment of or admission to the position of (associate) professor) and about salary increases.

GOALS	PERCENTAGE
Presence of possibilities to develop skills	90.7% (n = 136)
Evaluation of output	81.3% (n = 122)
Reflection upon leadership supervision	78.7% (n = 118)
Yearly appointments	71.3% (n = 107)
Maintenance/development of relations	59.3% (n = 89)

Table 20: Goals of job evaluations[58]. One research leader (1.3%) did not respond to this question.

Output indicators such as the number of publications (including the first authorship/co-authorship ratio), number of citations and the amount of research funding obtained are used by most of the research leaders interviewed in evaluating their staff. Other criteria reported in the interview study are: embeddedness in international networks, participation in internal work meetings and brainstorm sessions, style of leadership (if applicable), and educational contribution.

5.1.2 Pre-evaluation of research proposals

Most of the research proposals written by medical and health researchers in the Netherlands are pre-evaluated internally before being submitted to external funding agencies. The survey study shows that 91% of the medical and health research leaders participating in the study indicate that proposals are pre-evaluated on at least one level.

A. Organisational level: University

Research proposals of half of the responding research groups (53.0%) are subject to pre-evaluation organised by their university. Most of these research groups reports that researchers employed by their own faculty or institute – but not employed by their own research group – are asked to review research proposals (see Table 21). In 30.6% of these research groups research proposals are evaluated on this level by close colleagues. External researchers – who are employed outside the 'own' university – are not often approached in the evaluation procedure. Foreign researchers are rarely asked to review proposals.

PEERS	PERCENTAGE
Researchers employed in own faculty/institute	72.6% (n = 45)
Close colleagues	30.6% (n = 19)
Researchers employed elsewhere in the Netherlands	14.5% (n = 9)
Foreign researchers	8.1% (n = 5)

Table 21: Types of peers who pre-evaluate research proposals organised by universities

B. Organisational level: Medical faculty

The interview study shows that four medical faculties organise internal pre-evaluation of research proposals, depending on both the type of proposal and the funding agency (see Table 22). In one medical faculty, the research project, programme and equipment proposals are internally pre-evaluated by a research steering group before being submitted to KNAW or NWO. In another medical faculty, only research proposals submitted to the Dutch Cancer Society are pre-evaluated at this level. An evaluation committee that organises and evaluates these proposals has recently been established. Another medical faculty has set up an extensive internal evaluation procedure. In this medical faculty, three subcommittees of professors (from different disciplines) pre-evaluate personal research proposals (AGIKO[59], KNAW fellows or NWO Aspasia[60]), equipment proposals (NWO) and health care efficiency research proposals. Subcommittees are chaired by a non-university research institute manager. Comments, suggestions and recommendations are sent to the applicants and the Board of Directors. Finally, research proposals written by medical and health researchers employed by another medical faculty to obtain KNAW fellowship grants, AGIKO scholarships, NWO individual support for research and NWO financial resources for major investments in large scientific research equipment are pre-evaluated. A research committee provides information to the dean on the quality of the proposal and funding opportunities. It is the dean who finally decides whether the research proposal may be submitted.

Three medical faculties do not organise -at an institutional level- pre-evaluations of research proposals. This does not mean that research proposals are not reviewed before they are submitted to external funding agencies. For example, pre-evaluation of research proposals is organised by research groups or research divisions themselves, as indicated by one of the research managers interviewed.

[58] It should be stressed at this point that respondents are allowed to tick all suitable answers on questions about evaluation goals, peers and evaluation criteria. Therefore the sums of the percentages in table 2-5, 9-10, 14 and 15 are not equal to 100%.

[59] AGIKO grants are allocated in order to stimulate (young) doctors –who are training to become specialists - to combine their clinical tasks with scientific PhD research activities.

[60] The allocation of ASPASIA grant is intended to stimulate the promotion of female assistant professors to associate professor positions.

EVALUATION TYPE	NUMBER OF FACULTIES
Pre-evaluation of external research proposals organised by 'central' committee	3
Pre-evaluation of external research proposals organised by various subcommittees	1
No pre-evaluation of external research proposals	4

Table 22: Pre-evaluation of external research proposals internally organised by medical faculties

In the survey study, 70.5% (n = 86) of the responding research groups indicated that their research proposals are pre-evaluated at the faculty level. As already observed in the interview study with research managers, internal peers are often asked to evaluate research proposals (see Table 23). External peers (including international researchers) are rarely approached in the evaluation procedure. On average, external reviewers are invited in 2% of the research groups processing research proposal evaluations on this level.

PEERS	PERCENTAGE
Researchers employed by own faculty/institute	82.6% (n = 71)
Close colleagues	40.7% (n = 35)
Researchers employed elsewhere in the Netherlands	2.4% (n = 3)
Foreign researchers	1.6% (n = 2)

Table 23: Types of peers who pre-evaluate research proposals organised by medical faculties

C. Organisational level: Non-university research institute

According to the research managers, the four non-university research institutes included in the study (Netherlands Cancer Institute, Sanquin Blood Supply Foundation, Netherlands Institute for Health Services Research and TNO Prevention and Health) all organise pre-evaluations of research proposals. These qualitative observations contrast in part with the results of the survey study. 81.3% (n = 26) of the research group leaders affiliated with these non-university research institutes indicate that their research proposals are pre-evaluated at the institute level. In each of these research institutes, research proposals are evaluated regardless of potential external funding agency. In both the interview study and the survey study it was found that internal peers (usually two) are often asked to evaluate research proposals (see Table 24). A few research groups only approach external peers, including foreign researchers. In one of these four research institutes also customers are invited to peer review proposals.

PEERS	PERCENTAGE
Researchers employed by own faculty/institute	80.8% (n = 21)
Close colleagues	21.2% (n = 7)
Researchers employed elsewhere in the Netherlands	12.1% (n = 4)
Foreign researchers	9.1% (n = 3)

Table 24: Types of peers who pre-evaluate research proposals organised by non-university research institutes

The management team and/or research board of the research institutes plays an important role in the evaluation procedure. In the following interview quotes four research managers emphasise this: (1)*'Research proposals are presented to the research board (RB). The RB sends research proposals to two internal peers who evaluate the proposal. Suggestions and recommendations for improvement are sent to applicants, the scientific management team and the RB. The RB discusses the peer comments of proposals and sets up an advice (submission yes/no), which will be sent to the scientific management. After correction, it is the scientific management team who finally decides if researchers are allowed to submit the proposal to an external funding agency'*. (2)*'… budget and planning of research proposals are evaluated by the management team…'*. (3)*'…first ideas about research proposals are discussed in the management team…'*. (4)*'Results and outcomes of the peer review process are discussed in research meetings, which are attended by all researchers. The chair formulates conclusions'*.

D. Organisational level: Research group

The survey study results show that most research group leaders (83.9%, *n* = 135) internally organise the pre-evaluation of research proposals. Interestingly, some research leaders who do not organise pre-evaluations of proposals on a group level indicate that there is no need to organise such evaluations because they are experiencing an extensive evaluation procedure (and pressure) coming from non-university research institutes in which the groups are embedded. Furthermore, the evaluations organised by the non-university research institutes have a high impact: *'It is not allowed to submit proposals to external agencies without permission of two internal peers. A signature of both peers is required'. 'Researchers whose proposals receive a negative evaluation are strongly recommended not to submit their proposals to external agencies'.*

Both survey study and interview study show that pre-evaluation of research proposals organised by the groups themselves is often done by close colleagues employed in the same research group. Examples of close colleagues are research staff and senior researchers. Interestingly, from the qualitative interviews it appears that, in two research groups, evaluation committees are internally established. These committees are responsible for both the organisation and the peer review process of the pre-evaluation of research proposals. Table 25 shows that 81.5% (*n* = 110) of the research leaders who

indicate that they organise pre-evaluation of proposals approach direct colleagues to review the proposals. This is in contrast to the pre-evaluations organised by universities, institutes and medical faculties, which approach more often peers employed by the 'own faculty or institute'. External peers are of minor importance, in accordance with the description of the evaluation procedure of proposals on different levels as presented in preceding sections. In the interview study it was found that only one research group invites external peers on a regular basis (regardless of type of research proposal) in the review process. Another research group invites external peers only in the pre-evaluation process of research programmes.

	PERCENTAGE
PEERS	
Close colleagues	81.5% (n = 110)
Researchers employed by own faculty/institute	51.9% (n = 70)
Researchers employed elsewhere in the Netherlands	8.1% (n = 11)
Foreign researchers	9.0% (n = 12)

Table 25: Types of peers who pre-evaluate research proposals organised by research groups

The interview study shows that the research quality of proposals -indicated by peers- seems to be the most important criterion used in the pre-evaluation procedure at a group level. In two research groups, the addition of a second criterion, the fit of the proposal within the research programme or research priorities, extends the pre-evaluation procedure. Peers give general as well as more detailed comments and recommendations to improve the research proposal. They also estimate the chance of success of the proposal. The outcomes of pre-evaluations are intended for the quality assurance of research groups. In most groups it is up to the applicants to decide whether the outcomes should be incorporated. Nevertheless, some respondents stress the importance of evaluation outcomes. The following quote serves as an illustration: *'Researchers who have written a proposal that is evaluated negatively are strongly advised against submission to external agencies'.*

Three research leaders emphasise the high impact of pre-evaluations of proposals. In these groups, researchers are only permitted to submit the proposal externally after approval by the peers approached. In the survey study, research leaders were also asked their opinion on statements about pre-evaluations of research proposals. Table 26 shows that more than half of the respondents have the opinion that researchers should not be obliged to incorporate the comments and suggestions given by peers. Only a quarter of responding researcher leaders consider the incorporation of the outcomes of peer review compulsory. These results are in line with the results of qualitative interviews as presented above. According to almost half of the respondents, the final decision about submission or retraction of research proposals to external funding agencies should be left to the leader(s) of the research group. Almost a quarter of the respondents state that these decisions should be made by the researchers themselves. Finally, opinions differ greatly on the question of whether medical faculties or internal

non-university research institutes should have the authority to select and submit research proposals. A few researchers partly support this view. Just less than half of the responding research leaders disagree with this statement (see Table 26).

	DISAGREE FULLY	DISAGREE PARTLY	NEITHER AGREE NOR DISAGREE	AGREE PARTLY	AGREE FULLY
Researchers should be obliged to incorporate comments and suggestions from peers	13.5%	43.3%	19.1%	24.1%	0%
Research leader decides whether external submission of research proposals will occur	5.1%	19.1%	27.4%	33.8%	14.6%
Faculties or non-university research institutes should have the authority to select and submit research proposals	20.4%	26.1%	26.8%	24.8%	1.9%

Table 26: Opinions of research leaders on statements about pre-evaluations of research proposals, measured on 5-point scale

5.1.3 Research output evaluations

Output produced by medical and health research groups in the Netherlands is internally evaluated. In the same way as proposals, output is also evaluated at different levels. The evaluation procedure can, for example, be organised by universities, non-university research institutes, medical faculties and/or by research groups themselves. 98.8% (n = 158) of research leaders participating in the survey study report an evaluation of the output of their group at one level at least.

A. Organisational level: University

Of the 127 respondents, 85 (66.9%) are subject to research output evaluations organised by their university. In about half of these groups research output is evaluated at the highest level once in three to five years. Table 27 shows that the research output of more than one third of respondents is evaluated even more frequently.

A clear difference between the evaluation of proposals and the evaluation of output organised at the highest organisational level can be found with regard to involvement of peers. Table 27 shows that external peers – researchers not employed in the same medical faculty, institute or research group as the respondents – are asked by one third of the groups to evaluate their research output. In the evaluation of research proposals external peers are less often involved (see Table 21).

Box 5: Evaluation of research proposals and PhD students

In the Netherlands, PhD students are employed by universities. They are appointed for four years as an 'assistant in training' (AIO) and have a distinct academic position[61]. The survey study shows that medical and health research PhD students account for one third of the research group members. The ratio of PhD students and senior staff members in pre-clinical research groups is 1:1.25, in clinical research groups 1: 1.1 and in para-clinical research groups 1:0.8.

Qualitative interviews with research managers and research leaders show that a number of faculties and non-university research institutes have developed separate evaluation procedures to stimulate both PhD research and PhD students[62]. First, evaluation procedures have been developed to allocate internal (first-flow) money to PhD projects. PhD proposals written by senior researchers are evaluated yearly on the faculty level. Scientific quality is the main indicator. Also the research theme of the PhD proposal is important. According to a research manager, the success rate of PhD proposals increases if the proposed research theme is part of the faculty profile. Second, on a regular basis (mainly once a year), job evaluations are organised for PhD students. In some medical faculties and non-university research institutes job evaluations are organised and carried out by supervisors, while in other medical faculties and non-university research institutes PhD students are evaluated by an internal committee. Also the complexity of the evaluation procedure differs. For instance, one non-university research institute established a committee for each PhD student, which consisted of tutors/co-examiners and two senior researchers who were not members of the research group. The criteria are research planning, time schedule, publications, conference attendance, research resources, research training, teaching courses and content of the thesis. The outcomes of the job evaluations are reported by a coordinator. Negative outcomes are discussed with the head of the non-university research institute. According to this research manager, PhD students as well as members of the evaluation committee are satisfied with this evaluation procedure. In one of the medical faculties a PhD committee has been established in each of the six research sub-institutes. These PhD committees evaluate research progress and list the problems with supervisors experienced by PhD students. Third, a few research managers report that '*PhD students are stimulated during the whole research process'*. For example, PhD students are coached during the writing process of scientific articles. As one research manager explains further: '*The structure of the article, statistical data analysis and the detailed content of paragraphs are discussed. The research leader makes comments on the draft article. The article is under discussion three times at most. The PhD student loses the first authorship if the research leader rejects the article after the third discussion'.*

[61] The objective of this AIO system, which was introduced in 1985, is to provide advanced research training by way of active participation in university research and, to a limited extent, in teaching and administration. For more detailed information about the history of doctoral education policy in the Netherlands, see the thesis of Bartelse (1999), chapter 6.

[62] This (second) part of the interviews with research managers was unstructured. Due to this open character the questions about evaluation procedures that are internally organised to stimulate PhD students were not asked to every interview partner. Thus, no general conclusions can be formulated.

FREQUENCY	PERCENTAGE
Every 3–5 years	51.7% (n = 44)
Every 2–3 years	9.4% (n = 8)
Once a year	27.0% (n = 23)

Table 27: Frequency of research output evaluations organised by universities. Ten research leaders (11.7%) did not respond to this question.

The opposite is found for involvement of peers employed in the same medical faculty or non-university research institutes as the respondents. Table 89 shows that 43.5% of the respondents subject to output evaluations at this level indicate that researchers employed in their own faculty or institute, but not in their own research group, are asked to evaluate research output. As is shown in Table 21, involvement of these peers in the pre-evaluation process of research proposals on this level is much higher.

PEERS	PERCENTAGE
Close colleagues	57.6% (n = 49)
Researchers employed by own faculty/institute	43.5% (n = 37)
Researchers employed elsewhere in the Netherlands	29.4% (n = 25)
Foreign researchers	4.7% (n = 4)

Table 28: Types of peers involved in evaluation of research output organised by universities

Table 29 shows that publications (e.g. articles, reviews) are the most important items of research output to be assessed. In most cases the number of publications is used in the evaluation protocol as an indicator of research output. Also the impact factors of SCI scientific journals in which researchers have published is frequently used as a research output indicator. Furthermore, 36.5% of the research groups indicate that a citation analysis is part of the output evaluation procedure. In addition to the number of publications, it is also important to measure the number of dissertations. According to 67.1% of the research groups, the number of dissertations is used as an indicator in output evaluations. The amount of external research funding obtained is a quality indicator in almost half of the research groups. For most universities and non-university research institutes it is not important to compare the research output (inter)nationally. Finally, also the number of presentations or lectures given is of minor importance.

CRITERIA	PERCENTAGE
Number of publications	75.3% (n = 64)
SCI impact factors	68.2% (n = 58)
Number of dissertations	67.1% (n = 57)
External research funding	42.4% (n = 36)
Citation analysis	36.5% (n = 31)
Benchmarking	16.5% (n = 14)
Number of presentations	4.7% (n = 4)

Table 29: Criteria used in research output evaluations organised by universities

B. Organisational level: Medical faculty

All research managers but one report in the interview study that internal evaluations in which the research output of medical and health research is evaluated are organised at the faculty level. This finding is supported by the results of the survey study. According to 107 research groups leaders (84.3% of the respondents), research output evaluations are, on a regular basis, organised by their medical faculty. Table 30 shows that in about half of these groups (45.8%) these research output evaluations are organised yearly.

FREQUENCY	PERCENTAGE
Every 3–5 years	36.4% (n = 39)
Every 2–3 years	14.0% (n = 15)
Once a year	45.8% (n = 49)

Table 30: Frequency of research output evaluations organised by medical faculties. Four research leaders (3.7%) did not respond to this question.

Table 31 shows that, roughly speaking, three types of internal output evaluation procedures are used in different medical faculties: (1) no evaluation, (2) a restricted evaluation, and (3) an extensive evaluation. Only one medical faculty does not organise research output evaluations. A restricted evaluation of research programmes is organised by four medical faculties. The procedure consists of a (yearly) evaluation in which research output is broadly evaluated by three criteria: (1) external research funding, (2) number of publications and (3) impact of publication by using impact factors or other marks[63]. In one medical faculty also indicators of esteem (e.g. awards, invited lectures, conferences or symposia organised by researchers) are also taken into account. Internal output evaluations are coordinated by the research sub-institutes or on faculty level by a steering committee.

[63] An example to illustrate these publication marks is given here. A medical subfield is divided – on the basis of impact factors of SCI journals – into four (equal) parts. A publication in the first part of this subfield yields 4 points, a publication in the third part yield 3 points, etc. Furthermore, an interdisciplinary list of journals is formulated that consists of the most prestigious journals (Nature, Science, Cell, etc.). A bonus of 4 points is given for publishing in this category of journals.

Extensive internal output evaluations are organised by two medical faculties. The extensive internal evaluation procedures consist of: (1) a yearly evaluation in which research output is broadly evaluated and (2) a more comprehensive research output evaluation which is organised every three to five years. The broad analysis of the research output in these faculties is mainly made by counting the number of publications and citations, and by using impact factors. In one medical faculty this procedure was recently extended by taking parameters of esteem into account. Examples are editorial activities (chief editor or members of the editorial board) for journals listed in the ISI Journal Citation Reports and contributions to various national and international scientific councils, review committees and organisational committees for international congresses. In both medical faculties the extensive research output evaluation is organised and executed by internally established research advisory committees. These extensive evaluations are based on output and input figures which are mainly generated per research programme. Members of the internal committees evaluate the output of research programmes by criteria concerning publications (number, impact factor, citations, ratio of number of first authorships to number of co-authorships), dissertations, external funding (amount and the success rate in obtaining external research funding). A quantitative bibliometric analysis is sometimes part of the internal evaluation procedure. One faculty reports that the number of patents is used as an additional output criterion. The internal committees are given evaluation scores for each research programme evaluated. Both medical faculties use different scoring methods. In one faculty, four categories are used. A *I notation* is given to programmes of the international level; a *VI notation* is given to programmes that have the international level but are falling behind a little; a *VN notation* is given to programmes that have not achieved the international level yet but are progressing; an *N notation* is given to programmes that do not achieve the international level. Sometimes programmes are not evaluated because of their short duration; these programmes receive an *O notation (not evaluated)*. In the other faculty, the outcomes of research programmes are expressed on a four-point scale ranging from excellent to unsatisfactory. In addition, one of the medical faculties performed in 2001, as an experiment, an internal output evaluation research in which research output during the years 1999–2000 was evaluated by external peers. One Dutch scientist and one foreign scientist were selected by the dean of the medical faculty to peer-review research. In both faculties, preliminary evaluation outcomes are presented for comments to the programme leaders and heads of department who contributed to the programmes. On the basis of discussions it is possible to re-adjust the evaluation outcomes.

In general, final outcomes and recommendations of restricted as well as extensive output evaluations are presented to the faculty board. In a few faculties an advice drawn by the internal evaluation committee is added. Then the final outcomes are discussed, for example between the medical faculty board and the board of directors of the Academic Hospital. In two medical faculties, the final internal evaluation outcomes are compared with the outcomes of external output evaluations (for example,

DAG scores). Again, this task is accomplished by (a different) internal committee; for instance, one faculty established the committee 'Quality of Research'. In one faculty the results of this so-called strength-weakness analyses are expressed on a 4-point scale, ranging from strong to weak. The research programmes in this faculty are finally ranked (see Figure 8). This ranking is based on both the outcomes of the internal programme evaluation and the strength-weakness analysis.

1. Programme determines the profile of the faculty
2. After revision the programme will probably fit into the profile
3. Wait, maybe the programme will fit into the profile
4. Wait, the programme will stay outside the profile
5. Reorganise the programme
6. Stop the programme
7. No advice: programme is still reorganising

Figure 8: Ranking of research programmes

EVALUATION TYPE	**NUMBER OF FACULTIES**
No output evaluation	1
Restricted output evaluation: Step 1: Annual broad evaluation of research output Step 2: Presentation of evaluation outcomes to faculty board + discussion	4
Extensive output evaluation: Step 1: Annual broad evaluation of research output Step 2: Comprehensive research evaluation based on input and output figures once every 3–5 years Step 3: Presentation of evaluation outcomes to faculty board + discussion Step 4: Strength-weakness analysis	2

Table 31: Summarised evaluation of research output organised by medical faculties

C. Organisational level: Non-university research institute

The interview study with research managers shows that non-university research institutes also organise output evaluations[64]. A restricted output evaluation is organised by one non-university research institute. Two non-university research institutes organise more extensive output evaluations. In both institutes external peers, Dutch as well as foreign researchers, evaluate research output. In one of the institutes these external peers actually visit the researchers, as explained further in this quote: *'A two-day site visit is organized in which the peers are meeting the staff. During this visit posters are*

[64] The output evaluation of one of the non-university research institutes is not taken into account because their research output is not internally evaluated by a separate procedure, but as part of the 'knowledge audit' that is organised every 3–4 years. National and international peers evaluate, by the use of publication marks, the national and international knowledge position of this non-university research institute.

presented and oral presentations are given. The peer review committee closes the site visit by presenting a concluding evaluation report'. The criteria used in restricted and extensive output evaluations are comparable with criteria used in output evaluation organised by medical faculties (see above). However, one non-university research institute already takes societal impact of medical and health research into account. Another has recently performed a pilot study.

In general, final evaluation outcomes and recommendations are presented to the board of the non-university research institutes. On the basis of evaluation outcomes, two institutes formulate policy consequences which are written down, for example, in a list of improvements or a management letter.

D. Organisational Level: Research group

The survey study shows that most of the research group leaders (86.3%, *n* = 139) who participated in this study internally organise evaluations of research output. Table 32 shows that the frequency of output evaluations is high: two thirds of these research groups organise annual output evaluations.

FREQUENCY	PERCENTAGE
Every 3–5 years	18.0% (n = 25)
Every 2–3 years	8.6% (n = 12)
Once a year	67.6% (n = 94)

Table 32: Frequency of research output evaluations organised by research groups. 8 research leaders (5.8%) did not respond to this question.

Most research groups ask researchers employed in their own faculty or non-university research institute to evaluate their research output (see Table 33). The involvement of close colleagues in these output evaluations is lower than in output evaluations organised on a higher organisational level (56.8% and 70.5% respectively). Interestingly, 21 research leaders (15.3%) indicate that they ask foreign researchers to evaluate the research output of their group. Compared to the low proportion of research leaders who indicate that foreign researchers are involved in the output evaluation procedures organised at a higher organisational level (ranging from 3.8% to 9.1%, see previous sections), the rate of 15.3% is high.

PEERS	PERCENTAGE
Researchers employed by own faculty/institute	78.1% (n = 107)
Close colleagues	42.6% (n = 58)
Researchers employed in the Netherlands	22.1% (n = 30)
Foreign researchers	15.3% (n = 21)

Table 33: Type of peers who evaluate research output organised by research groups. Three research leaders (2.1%) did not respond to this question.

As is shown in Table 34, most of the research groups use the number of publications and their SCI impact as indicators of research quality. Not only are published papers important, but the number of publications that are still in the pipeline is also an output indicator. In the interview study it was found that some research groups make an inventory of the number of articles that are in press, accepted, submitted, or in preparation. In addition, a few groups report that they evaluate the future prospects by taking work in progress into account. Table 35 shows that the amount of research funding obtained from external agencies is an output indicator in almost two thirds of the research groups. Research groups use this indicator more often than universities or non-university research institutes to measure research output (60.4% compared to 44.1% and 56.2%; see Table 28). Also, the number of dissertations is an important output indicator, being used by some two thirds of the research groups. One third of the research groups use a citation analysis to evaluate their output. The number of presentations are important in almost 20% of the research groups. Furthermore, in the qualitative approach it was found that some groups make a distinction between invited lectures and unsolicited lectures given by their researchers during conferences. Table 34 shows that most research groups do not compare their research output with that of similar (inter)national research groups.

INDICATOR	PERCENTAGE
Number of publications	88.4% (n = 123)
SCI impact factors	82.0% (n = 114)
Number of dissertation	61.2 % (n = 85)
External research funding	60.4% (n = 84)
Citation analysis	31.7% (n = 44)
Number of presentations	19.4% (n = 27)
Benchmarking	13.7% (n = 19)

Table 34: Criteria used in research output evaluations organised by research groups. Two research leaders (1.4%) did not respond to this question.

In qualitative interviews research leaders mention four additional output indicators used in internal evaluations organised by research groups:

- Number of abstracts

- Number of published books
- Number of conferences or symposia organised by researchers
- Participation of researchers in committees (editorial boards of scientific journals, evaluation and audit committees)

5.2 Implications of internally organised research output evaluations

Research evaluation is an indispensable tool in improving research. Internal quality assurance of research input, throughput and output depends on the efforts of research leaders and staff. Therefore, a great deal of importance is attached to the use and implementation of evaluation outcomes. This section on the effects of internal research output evaluations is based on open unstructured interviews with research managers and research leaders. Some qualitative observations are supported by quantitative results of the survey.

5.2.1 Implications perceived by research managers

Research managers were asked to describe whether and how the outcomes and recommendations of output evaluations organised by medical faculties and non-university research institutes are implemented. They were also asked to assess the implications of output evaluations organised by medical and non-university research institutes.

In most medical faculties and non-university research institutes the evaluation outcomes and recommendations have had **direct implications** for the local research policy[65]. Five different direct implications are reported by research managers:

(a) Re-adjustment of research programmes. For example, a negative evaluation outcome can result in a phase-out of research programmes (*'…during the past 10 years, 15 to 20 research programmes were finished...')* or research groups are recommended to merge *('…the merger of a small research group and a big research group has enhanced the research quality as measured in a later evaluation...').* On the other hand a positive evaluation can result in a development of new research programmes.

(b) Contribution to the formulation (and adjustments) of the total research strategy and the profile of medical faculties or non-university research institutes. One example: *'On the basis of*

[65] It should be noted that there may be reasons, for example when a new research leader has just been appointed or when the research field concerned is moving rapidly, that make the board or dean decide not to implement the recommendations.

outcomes of internal output evaluations, immunology, haematology and organ transplantation became the main areas that should be stimulated within our faculty'.

(c) Redistribution of financial resources. In two medical faculties the evaluation results are used to reallocate resources. This implies that the amount of money that is internally allocated (first-flow money) to research departments and research programmes is variable. As one research manager explains in more detail: '*25% of the structural first budget source of scientific departments*[66] *depends on the results of the internal output evaluation. Half of this 25% is allocated on the basis of two criteria: amount of external funding (second-flow money) obtained and the number of theses. The other half of 25% is allocated on the basis of the number of publications in leading journals (defined as journals listed in the first part of the SCI index)* [67]*'*. This allocation of research money on the basis of performance aims to gain some flexibility in research budget and to stimulate research quality.

(d) Appointment of staff. The following quotes of two research managers from non-university research institutes serve as examples: *'In department X, supervision of PhD students was evaluated as being inadequate. The number of PhD students was too large in proportion to the number of staff. The research leader responded by writing a job description for a second head of the department. This second head will supervise the PhDs. At this moment the position is still a vacancy'. 'In 1990, the faculty appointed (20 FTE) research staff to develop and to conduct health care research'*

Two research managers pointed out that the internally organised output evaluations only have **indirect implications** for local research policy. One research manager argues that the advice, which is formulated in reply to the evaluation outcomes, is not binding and intended for internal quality assurance of university research institutes: *'It is up to the institutes to decide whether or not the outcomes and recommendations are incorporated'*. The other research manager explained that the dean offers research leaders the opportunity to clarify the evaluation outcomes and recommendations. Recommendations are implemented only if total agreement is reached.

5.2.2 Implications perceived by research leaders

Research leaders were asked for their opinion on the usefulness of results and recommendations of internal output evaluations organised by universities, research institutes and non-university research

[66] The first experiences with this approach will be evaluated. A positive evaluation will increase this percentage to maximum of 50%.
[67] A list of leading journals is made for each medical subfield.

institutes. In addition, they were asked to indicate the impact of internal evaluation outcomes. The outcomes and recommendations of internal output evaluations are experienced as (very) useful by more than half of the responding research leaders. Table 35 shows that only a few research leaders (*n* = 14, 9.5%) believe that these evaluations are not (very) useful for their own research group. Some of them make the criticism that the evaluation criteria used do not allow for differences among medical disciplines. According to half of the respondents the results and recommendations of internal output evaluations are generally implemented intensively in the research policy of their groups. Only a small proportion (12.3%, *n* = 18) of the responding research leaders indicate that they did not (or only very rarely) implement outcomes and recommendations (see Table 35). One of the reasons given is the lack of evaluation consequences. As one of the research leaders points out: *'The outcomes are only of importance if they actually influence the extent/size of the research budget. This is currently not the case'*. The extent to which evaluation outcomes and recommendations have already been implemented in the research policy of groups is often used as an indicator in the ensuing output evaluations. According to 12.2% of the research leaders a follow-up is a standard indicator in the evaluation procedure while 60.9% of the respondents indicate that a follow-up is sometimes conducted and a quarter of the research leaders never noticed any follow-up activities.

EFFECTS PERCEIVED BY RESEARCH LEADERS	**PERCENTAGES**		
Experiences outcomes and recommendations	57.4% (very) useful	33.1% neutral	9.5% not useful
Implementation of outcomes and recommendations	52% (very) intensive	35.6% neutral	12.3% not intensive

Table 35: Implications of output evaluations that are internally organised by universities, research institutes and non-university research institutes as perceived by research leaders: a summary

5.2.3 Future prospects of research output procedures developed by medical faculties and non-university research institutes

The interview study with research managers in 2000-2001 revealed that there is a great deal of internal discussion about the evaluation procedures used by medical faculties and non-university research institutes. Four research managers report that the whole internal evaluation procedure will undergo changes in the near future (Table 36). In one medical faculty, research output is not being evaluated yet. In the near future, research output indicators will be developed on a faculty level. Another research manager stated that new evaluation guidelines are being developed in the faculty, but that it is (still) not clear in how much detail the output of research will be evaluated. One non-university

research institute seriously considers adopting a research output procedure that is currently used by an inter-university research institute. Finally, the board of another non-university research institute would like to develop and introduce a model that would steer medical and health research in the whole organisation. Scientific research quality (indicators will be borrowed from the DAG protocol) and research management should be the most important criteria.

In two medical faculties and one non-university research institute a few small changes in the research output evaluation procedure will be implemented (Table 36). According to one research manager, a site visit will introduce a new item in the internal procedure used by the faculty to evaluate research output. In another medical faculty, a steering group is just discussing the possibilities of involving external peers in research output evaluations. Furthermore, this faculty is developing a publication scoring system ('*number of points depends on type of publication*') and is discussing the extension of the internal output evaluation procedure by incorporating citation analysis ('*probably this will only be applied to top-class researchers*'). Finally, the research manager from the non-university research institute reported that the institute is thinking about the development of societal quality indicator(s) and its incorporation into the internal output evaluation procedure. This institute is engaged in applied policy research for policy makers in public healthcare. This research is conducted for a large number of clients, ranging from governmental bodies and (umbrella organisations of) healthcare insurers to professional groups, healthcare organisations and patient and consumer organisations. Therefore, not only scientific quality but also societal quality is essential.

CHANGES IN RESEARCH OUTPUT EVALUATION PROCEDURE	NUMBER OF FACULTIES	NUMBER OF NON-UNIVERSITY RES. INSTITUTES
Implementation of a whole new evaluation procedure	2	2
Implementation of small changes in evaluation procedure	2	1

Table 36: Changes in research output evaluation procedures in medical faculties and non-university institutes

5.3 Concluding remarks

In general, medical and health research is internally evaluated in different phases of the research process. According to the majority of the research leaders who participated in the survey study, both research input (jobs and proposals) and research output are evaluated (Table 37). These evaluations are organised by different organisational entities: universities, medical faculties, non-university institutes and research groups. The relationships of medical and health researchers with their internal environment, as found in the interview study and survey study, are presented in Figure 9. The

conclusion with respect to the features of evaluations procedures developed and used internally in order to assess staff, research proposals and research output are presented separately in subsequent sections.

Research evaluation	**Percentage**
Job evaluation	95%
Pre-evaluation research proposals	91%
Research output	98%

Table 37: Percentage of research leaders indicating that research evaluations are organised internally

Job evaluations

In conclusion, the interview study as well as the survey study shows that the majority of the research leaders evaluate staff members yearly. It appears that job evaluations serve multiple purposes. They are not only used to the assess staff but also to improve conditions in order to stimulate (individual) research quality. Furthermore, the results of job evaluation are used for the development of researchers' careers.

Pre-evaluation of research proposals

In sum, it appears from both the interview study and the survey study that research proposals written by medical and health researchers are pre-evaluated at different organisational levels before they are submitted to external funding agencies. At the medical faculty and non-university institute level these pre-evaluations are organised by steering committees. What proposals are pre-evaluated varies among faculties, depending generally on the type of proposal (personal or programme grant) and the funding agency. Non-university research institutes pre-evaluate any research proposal written, regardless of external funding agency. The procedures are comparable to those used by medical faculties, but are more detailed. Also, both the research board and the research management team are highly involved in the pre-evaluation. Research groups generally organise pre-evaluations of proposals. Research quality of proposals seems to be the main criterion. Peers give comments and suggestions to improve research proposals. In most of the research groups, such advice is not binding.

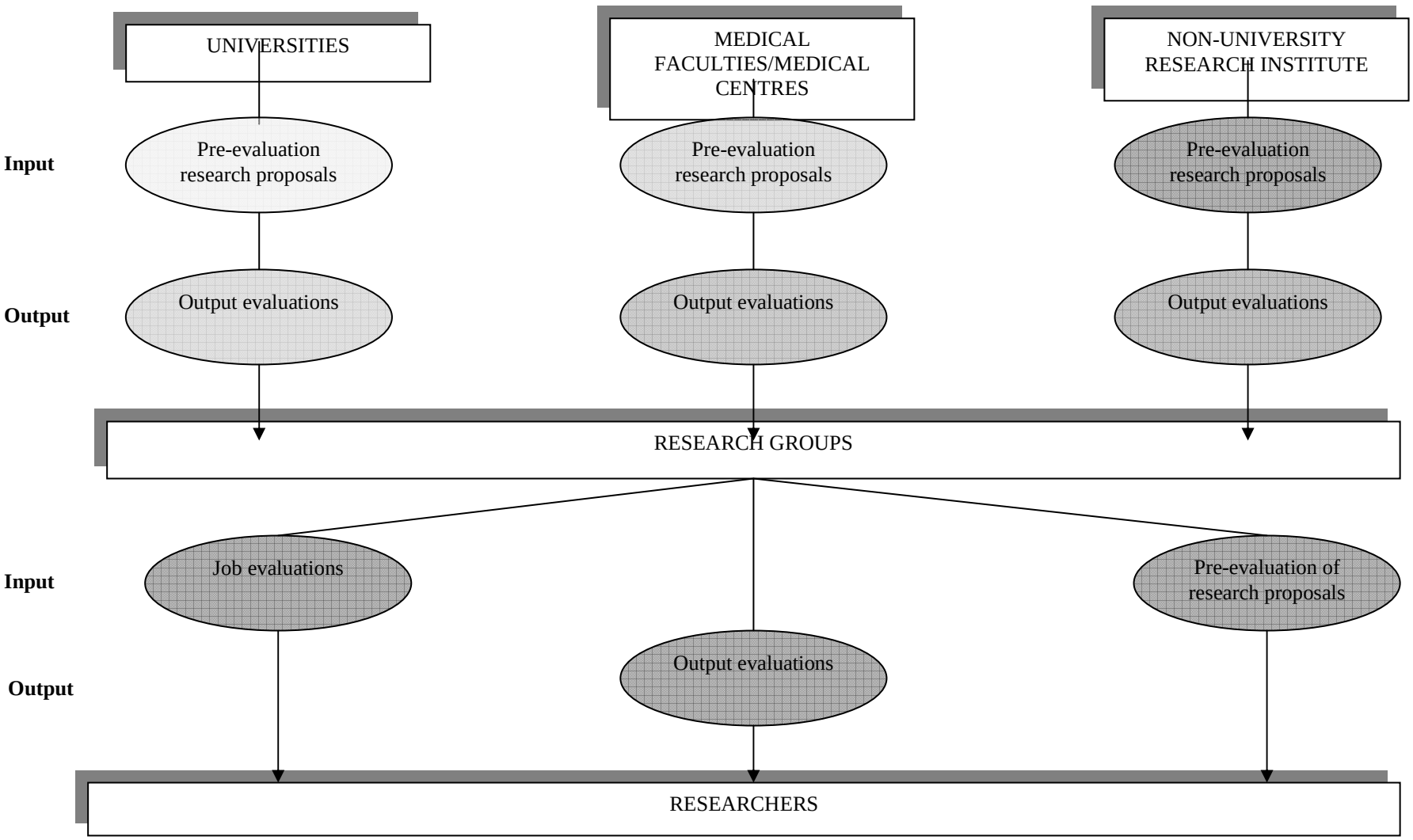

Figure 9: Relations between different internal actors organising evaluation practices

Finally, it can be concluded that internal peers are mainly involved in the pre-evaluations of research proposals organised at any organisational level. External peers are rarely involved.

Research output evaluations

To summarise, results of both the interview study and the survey study show that output evaluations are prevalent. However, the complexity of evaluation practices varies among medical faculties and non-university research institutes. Three types of output evaluation practices – none, restricted and extensive – can be distinguished. Figure 10 conceptualises these three types. The boundary lines between the different types are not clear or sharp. From the interviews with research managers it can be concluded that one evaluation type can gradually be converted into another over time. The empirical data show that the arrows should be drawn to the right. None of the medical faculties and non-universities that organise extensive research output evaluations are planning to switch to more restricted output evaluations types.

The findings of both the interview and survey study suggest that the involvement of external Dutch peers is (much) higher in internal output evaluations organised at any organisational level than their involvement in pre-evaluation of research proposals.

Finally, it can be concluded that the number of publications and their SCI impact are the most popular indicators in output evaluations organised by universities, medical faculties, non-university institutes and research groups alike.

Figure 10: Conceptualisation of the three types of research output evaluations organised by medical faculties and non-university research institutes

Implications of internally organised output evaluations

In sum, it appears that internally organised research evaluations are frequently used by the (faculty) board to re-adjust, develop and stimulate local research programmes. The (faculty) board also uses internal evaluations outcomes to formulate and adjust their research strategy and profile. The survey reveals that most research leaders take positive views on the outcomes of internal research evaluations. Most of the research groups implement the recommendations given by peers.

The question arises as to whether internal research evaluations and other research management items are positively correlated with research performance. In addition, research leaders were asked in the interview study whether (six) different research management items could influence (positively or

negatively) the performance of the research group. According to para-clinical, pre-clinical and clinical research leaders the staffing policy internally developed to employ research staff is a very important precondition to performing well (Table 38). Also development of a scientific strategy and collaboration with other (inter)national research groups seem to be important conditions. In choosing new research topics, apparently no attempts have been made to maintain consistency among research themes. The interview study shows that the connection with previous research is no prerequisite for high performance. A clear difference was found with the importance attached by research schools. The pre-clinical and clinical research group leaders interviewed stated that it is not important to be a member of a research school. The two para-clinical research group leaders take more positive views.

	PARA-CLINICAL GROUPS (N=2)	**PRE-CLINICAL GROUPS (N=4)**	**CLINICAL GROUPS (N=6)**
Staffing policy	++	++	++
Scientific strategy	+/-	+	+
Consistency of research themes	-	-	-
Member of research school	+/-	--	-
Collaboration	+	+	+

Table 38: The importance of research management items for research performance. ++ very important, + important, - hardly important, -- not important.

The interview study provides insufficient data to answer the question of whether research management items correlated positively with research performance. Therefore, a survey was developed for investigating this research question further. In the next chapter (Chapter 6) a series of bivariate and multivariate analyses are presented.

CHAPTER 6: RESEARCH MANAGEMENT AND PERFORMANCE

This chapter presents the main results of the survey study. The survey was sent in 2002 to research leaders of Dutch health research groups. In the first part of this chapter the response rate and the analysis of non-response are examined. In the second part comparisons are made between contingencies, views on research management, research management activities and performance. Data from para-clinical, pre-clinical and clinical research groups are presented separately. To investigate whether views about research management and research management activities are positively related to research performance, bivariate and multivariate analyses of contingencies, research management and performance are presented.

6.1 Data collection

6.1.1 Response rate

All 412 research leaders of Dutch academic health research groups (mainly chaired professors) were approached. A total of 31 research leaders were non-eligible. They turned out to be attached to their research group only for a short period of time (less than 6 months) or could not answer the questions because hardly any research was done in the group (focus on patient care) or had just left the research group (professors emeritus). The research group leaders returned 162 questionnaires, resulting in an individual response rate of 42.5% of the eligible population of 381. Two questionnaires could not be used because some parts of the survey were not filled in. Therefore, 160 questionnaires were analysed. Group members (junior and senior staff) were not involved in the study. Therefore, the data show the perspective of the research group leader.

6.1.2 Analysis of non-response

The response rate in the eight medical faculties varies from 24% to 60% (see Table 39 for an overview). Three research managers from medical faculties were not willing to write a letter of recommendation. The response rate of research leaders operating in faculties (44%, 34%, 24%) is, on average, lower than that of research leaders operating in faculties who did write a letter of recommendation (60%, 56%, 44%, 33%, 25%). In the faculties with the lowest response rates (below 30%) the principal investigator contacted the research leaders or the secretaries of the non-responding

groups to list the reasons for non-response. Involvement in reorganisation, lack of time and sabbatical leaves were most often mentioned as reasons for non-response.

	NUMBER OF GROUPS	RESPONSE RATE (%)
A	13	34
B	28	56
C	18	60
D	18	44
E	15	44
F	9	25
G	12	24
H	14	33
Total	**127**	**39.6%**

Table 39: Overview of the number of participating research groups in eight medical faculties (indicated with A until H). During this empirical study (in 2002), most of the medical faculties were co-operating to some degree with an academic medical centre (see also footnote 59 in Chapter 5).

The response rate in the four non-university health research institutes differed from 25% to 75% (Table 40). The institute with the lowest response rate was contacted. Both non-responding and responding research leaders indicated that there would be no differences in internal research policy between units because they agreed on most of these items.

	NUMBER OF GROUPS	RESPONSE RATE (%)
A	21	58
B	6	75
C	3	25
D	3	60
Total	**33**	**53.8%**

Table 40: Overview of the number of research groups in four non-university health research institutes (indicated with A until D).

Of the 160 respondents, 22 were research leaders of para-clinical research groups, 75 were research leaders of pre-clinical groups and 63 were research leaders of clinical research groups (see Table 41). Para-clinical research groups accounted for 15.4% of the total population and 13.8% of the response sample. 34.3% of the total population and 46.8% of the response sample were classified as leaders of pre-clinical research groups. Clinical research groups made up 49.7% of the total population and 39.4% of the sample of respondents. This study does not intend to draw general conclusions about the relation between management activities, contingencies and performance in the whole population of Dutch health research leaders. The classification of para-clinical, pre-clinical and clinical research is used in each analysis (as described in the next sections).

	TOTAL POPULATION	NON-RESPONDENTS	RESPONDENTS
Para-clinical research groups	N = 59 15.4% of total population	N = 37	N = 22 13.8% of response sample
Pre-clinical research groups	N = 131 34.3% of total population	N = 56	N = 75 46.8% of response sample
Clinical research groups	N = 180 49.7% of total population	N = 127	N = 63 39.4% of response sample
Unknown	N = 2 0.5% of total population	N = 2	-

Table 41: Approached and responding health research groups

The effect of non-response depends on the proportion of the population who do not respond and to the extent to which those not responding are biased, i.e. systematically different from the whole population (Fowler, 1993). Fowler argues that a generalisation that seems to hold for most mail surveys is that people who have a particular interest in the subject matter or the research itself are more likely to return mailed questionnaires than those who are less interested. This means that a survey may be biased significantly in the way it is related directly to the purpose of the research. In order to test whether there were significant differences between respondents and non-respondents dependent performance variables were compared. The performance of the research groups measured that responded was compared with the performance of the groups that did not respond.

1. Number of SCI papers

On the basis of data obtained from the ISI database, the numbers of papers (co-)authored by responding research group leaders that were published in international SCI scientific journals in 1999, 2000 and 2001 were compared with those of non-respondents. The mean difference between the two groups is very small (24.35 vs. 24.83 publications) and the variance within both groups is large (SD 24.49 and 24.82 respectively). Analysis of variance therefore shows a small and non-significant F-value. It can be concluded that there is no significant difference in mean number of SCI papers published by respondents and non-respondents (see Table 432.

2. NWO rating

In addition, by using the electronic database Delfi, the NWO rating of the approached research group leaders was measured (Table 42). NWO rating was expressed on a 4-point scale. Three rating indicators were used with regard to NWO grants, NWO peers and information about NWO funding programmes. Table 43 shows that there is no significant difference for this variable between respondents and non-respondents.

3. Number of submitted NWO research proposals

Finally, the number of submitted NWO research proposals received in competition for research grants from respondents were compared with those from non-respondents. A submitted research proposal was rated as valid if: (1) the name of the research group leader was mentioned in the research proposal as main applicant, fellow applicant, research leader or researcher and (2) the research proposal was entered in Delfi. Interestingly, those who participated in the study had submitted on average significantly (F=9.39**) more research proposals to NWO (mean 2.48) than those who did not participate (mean 1.63). This difference can be explained by the relatively high percentage of groups that did not submit any research proposal to NWO during the years 1999-2001. 43.8% of the non-respondents did not submit research proposals to NWO, compared to 33.5% of the respondents. Probably, these non-respondents are group leaders who are not so much focused on basic health research. It may also indicate that these group leaders, after estimating their chances for NWO success, think that their research does not have the required quality level and innovativeness. In addition, NWO success percentages are generally low. Especially clinical research groups often search for alternative research funding agencies, as for example charity funds.

	RESPONDENT N = 160	**NON-RESPONDENT N = 222**	**F VALUE**
Number of SCI papers	24.35 (24.49)	24.83 (24.82)	0.779
NWO rating	3.12 (1.15)	2.92 (1.20)	2.466
Number of submitted NWO research proposals	2.48 (3.38)	1.63 (2.39)	9.385**

Table 42: Analysis of non-response (analysis of variance)

In conclusion, both the response rate and the main part of the non-response analyses suggest that the participating medical and health research groups can be regarded as representative of medical health research in Dutch universities and non-university institutes. In the next section the results found for para-clinical, pre-clinical and clinical research groups are presented.

6.2 Results

6.2.1 General information about respondents

The average age of the respondents was 54 years (sd 6.4), and the majority were male (91%). On average, respondents had obtained their PhD about 20 years (sd 7.3) before this study was performed. 79% of the respondents were working in a university setting and 21% in a non-university research institute. Writing and publishing scientific articles (93%), development of new knowledge (58%) and

training of young researchers (48%) were the most important goals of the research groups supervised by the respondents.

6.2.2 Contingencies

Table 43 compares the contingencies in para-clinical, pre-clinical and clinical research groups. It shows that, on average, 19 research staff members were working in a research group (SD 16.8). More time was allocated to research (F = 3.93; P<0.05) in pre-clinical research groups. As could be expected, more time was allocated to patient care (F = 8.71; P<0.01) in clinical research groups than in pre- and para-clinical research groups, which spend almost no time on patient care. No differences were found in time allocated to education (mean 1.7 on a 5-point scale) and supervision of PhD students (mean 2.4) among para-clinical, pre-clinical and clinical research groups. The heads of the para-clinical research groups had less (F = 3.19; P<0.05) research experience (21.3 years on average) than heads of pre-clinical and clinical research groups (25.5 and 23.6 years on average respectively). There is a misunderstanding among respondents with regard to the definition of research management. When research leaders were asked about their managerial experience, 42% indicated that they do not have any management experience. Interestingly, these respondents do not see leadership as a managerial task. The groups differed with regard to foreign research experience (χ^2 = 9.84; P<0.01). Whereas leaders in pre-clinical and clinical research groups had been working in foreign countries for about two years, heads of para-clinical research groups generally lacked international experience[68].

[68] Foreign research experience was measured as a quantitative measurement level (ratio scale). This variable is not normally distributed; for example 39% of the leaders do not have any foreign experience at all. Therefore we recoded foreign experience into a dummy variable with categories: no foreign experience (coded 0) and foreign experience (coded 1) and used this dummy in regression analyses. Recoding into a variable with three categories (trichomy): no foreign experience (coded 0), 1-2 years foreign experience (coded 1) and > 2 years experience (coded 3) shows the same results compared to the dummy.

CONTINGENCIES	PARA-CLINICAL (N = 19)	PRE-CLINICAL (N = 73)	CLINICAL (N = 61)	F VALUE
Size (full-time equivalents, fte)				
Research staff	18.7 (21.1)	22.0 (19.8)	17.3 (11.2)	1.30
Time allocation (5-point scale)				
Research	3.1 (1.1)	3.8 (2.1)	2.9 (1.3)	3.93*
Education	2.1 (1.3)	1.7 (0.9)	1.5 (0.9)	2.44
Patient care	0.8 (1.5)	1.3 (1.5)	2.3 (1.9)	8.71**
Supervision of PhD students	2.4 (1.3)	2.6 (1.2)	2.3 (1.0)	1.04
Age (years)				
Research experience	21.3 (7.6)	25.5 (6.6)	23.6 (7.0)	3.19*
Management experience	7.1 (7.6)	5.7 (7.4)	5.7 (6.3)	0.30
Foreign experience	0.6 (1.2)	2.0 (3.2)	2.3 (4.5)	1.74
Supervising experience	12.5 (6.5)	13.0 (7.3)	11.2 (5.9)	1.18

Table 43: Comparison of contingencies in para-clinical, pre-clinical and clinical research groups

Table 44 shows that the staffing structure in para-clinical, pre-clinical and clinical groups is more or less the same. The leaders of research groups and senior staff together accounted for 48% (para-clinical groups), 34% (pre-clinical groups) and 32% (clinical groups) of total staff, whilst junior staff accounted for 38% (para-clinical groups), 31% (pre-clinical groups) and 41% (clinical groups) of total staff. Not surprisingly, the staffing structure of research groups significantly differs ($F = 8.37$, $P<0.01$; $\chi^2 = 31.67$; $P<0.01$). Laboratory technician staff in para-clinical groups accounted for only 4% of total staff (mean 0.6 fte), compared to 28% (mean 5.9 fte) and 19% (mean 2.9 fte) in pre-clinical and clinical research groups.

STAFFING STRUCTURE	PARA-CLINICAL (N = 19)	PRE-CLINICAL (N = 73)	CLINICAL (N = 61)	F VALUE
Research leader (fte)	1.4 (1.0)	1.1 (0.9)	1.2 (0.9)	0.41
Senior scientific staff (fte)	6.6 (10.1)	6.0 (6.4)	3.9 (3.0)	2.50
PhD students (fte)	6.4 (5.6)	6.4 (5.2)	6.4 (4.6)	0.00
Laboratory technician staff (fte)	0.6 (1.1)	5.9 (7.3)	2.9 (3.8)	8.37**
Technical support staff (fte)	1.8 (0.4)	1.5 (3.9)	1.3 (2.6)	1.28

Table 44: Comparison of staffing structure in para-clinical, pre-clinical and clinical research groups in fte

6.2.3 Research management

Tables 45 and 46 compare the views and judgements of para-clinical, pre-clinical and clinical research leaders about research management items. Furthermore, research management activities of research leaders in para-clinical, pre-clinical and clinical research groups are shown in these tables.

Internal control

Medical and health research groups have serious problems in attracting qualified scientific staff such as PhD students, senior staff and laboratory technicians. About half of the respondents (52%) indicate that they are short-handed. This seems to be the main explanation of the problems that exist in carrying out health research[69].

In general, medical and health research leaders spend 20–25% of their time on internal management activities. No differences between leaders of para-clinical, pre-clinical and clinical leaders were found.

In spite of these managerial tasks, research group leaders are, in general, still highly involved (mean 4.1 on a 5-point scale) in medical and health research undertaken within their own research group. Internal research meetings are frequently organised by research leaders. Differences between para-clinical, pre-clinical and clinical research groups are very small. Pre-clinical research groups organise, on average, more research meetings (mean 3.6, measured on a 5-point scale) than para-clinical (mean 3.4) and clinical research groups (mean 3.5).

Of the research leaders who responded, 80% organise meetings to discuss the long-term research policy of the group. Researchers participate in these discussions. In 96% of the research groups employees are asked to participate in these discussions. In addition, other researchers such as researchers employed in the same university or research institute (40%), researchers employed in the same research school (22%), researchers employed in a different university or research institute (11%) and foreign researchers (14%) are asked to participate in these discussions. No differences were found among the three types of medical and health research groups.

In clinical research groups 36% of the communication between leaders and their research staff members goes directly via e-mail, telephone and personal talks. This percentage is significantly (F = 5.75; P<0.01) higher than in pre-clinical (27.7%) research groups.

In general, medical and health research leaders often offer their staff opportunities to develop their research skills (mean 2.6) and to work flexible hours (mean 2.2). In order to motivate research staff, special commendations are sometimes given (mean 1.7). The mean use of offering financial incentives to staff differs significantly (χ^2 = 6.44; P<0.05) among research groups. Clinical research groups

[69] Research groups face problems and difficulties in conducting health research due to shortage of staff (52% of the respondents), equipment (33%), library services (28%), research materials (16%), data processing (12%) and laboratory room (6%). Compared to para-clinical and clinical research groups, pre-clinical research groups experience more deficiencies in laboratory room (χ^2 = 7.2*) and library services (χ^2 = 9.9**).

leaders less frequently give financial incentives (mean 1.3) than para-clinical (mean 1.5) and pre-clinical research group leaders (mean 1.5).

Not surprisingly, almost every medical and health research group (96%) uses internal job evaluations to evaluate their research staff members. 87% of the health research groups organise internal pre-evaluations of research proposals before submission to external research funding agencies will take place. Pre-evaluation of research proposals is not only organised by research groups but also by medical faculties (69%), universities and non-university research institutes (56%). On average, more clinical groups (80%) seem to experience pre-evaluation of research proposals organised by medical faculties than para-clinical and pre-clinical research groups (71% and 61% respectively; $\chi^2 = 5.5$, n.s.). Group leaders think that internally organised pre-evaluation of research proposals will result in an increased opportunity to receive an external research grant (mean score 3.8, measured on a 5-point scale). Moreover, 88% of the medical and health research groups evaluate their research output internally. Evaluation of research output is also organised internally by medical faculties (82%), universities and non-university research institutes (69%). On average, fewer para-clinical groups experience internal evaluation of research output organised by medical faculties (68%) than pre-clinical and clinical research groups (84% and 85% respectively). Internally organised research evaluations seem to be the most important type of evaluation for medical and health groups (mean importance 3.4).

INTERNAL CONTROL	**PARA-CLINICAL (N = 19)**	**PRE-CLINICAL (N = 73)**	**CLINICAL (N = 61)**	**F VALUE**
Resource control: personnel (5-point scale)	2.9 (1.4)	3.1 (1.0)	3.0 (0.9)	0.16
Time spent on internal management activities (6-point scale)	2.6 (1.3)	2.3 (1.2)	2.2 (1.0)	0.89
Research commitment (5-point scale)	4.1 (0.5)	4.1 (0.5)	4.1 (0.4)	0.04
Communication				
Research meetings (5-point scale)	3.4 (0.4)	3.6 (0.6)	3.5 (0.5)	1.32
Process of communication (percentage)	31.6 (12.7)	27.7 (15.0)	36.3 (14.2)	5.75**
Research policy planning (dichotomous)	90.0 (20.0)	80.0 (40.0)	80 (40)	1.48
Rewards (3-point scale)				
Development of research skills	2.6 (0.2)	2.7 (0.3)	2.6 (0.3)	0.60
Flexibility	2.4 (0.5)	2.2 (0.5)	2.1 (0.5)	2.05
Special commendations	1.5 (0.7)	1.8 (0.7)	1.7 (0.7)	0.79
Financial bonus system	1.5 (0.6)	1.5 (0.6)	1.3 (0.5)	2.94
Internal organisation of research evaluations (dichotomous)				
Job evaluation	100.0%	93.2%	98.4%	1.61
Pre-evaluation of research proposals:				
Level of research group/department	88.0% (33.0)	91.0% (40.0)	93.0% (25.0)	2.35
Level of faculty/research institute	71.0% (47.0)	61.0% (49.0)	80.0% (41.0)	2.82
Level of university/research institute	65.0% (49.0)	56.0% (50.0)	60.0% (49.0)	0.79
Evaluation of research output:				
Level of research group/department	84.0% (37.0)	86.0% (35.0)	93.0% (36.0)	1.10
Level of faculty/research institute	68.0% (48.0)	84.0% (37.0)	85.0% (36.0)	1.48
Level of university/research institute	58.0% (51.0)	73.0% (45.0)	74.0% (44.0)	0.95
Effectiveness of pre-evaluation of research proposals (5-point scale)	4.0 (0.5)	3.7 (1.1)	3.8 (0.9)	1.07
Importance of internally organised research evaluations (5-point scale)	3.5 (1.1)	3.4 (0.8)	3.5 (0.7)	0.21

Table 45: Comparison of internal control items in para-clinical, pre-clinical and clinical research groups (see 4.3.1 for a description of the measurement scales used)

External control

Research leaders spend an average of 20–25 days per year on external research activities such as giving lectures, attending or organising conferences, participation in editorial boards of journals, participation in audit committees and participation in assessment committees. Compared to the amount of time allocated to internal management activities, medical and health research leaders on average allocate less time to external management activities (mean 1.4).

Externally organised research evaluations such as DAG 1998 and the accreditation of research schools are considered less important than internally organised research evaluations. Medical and health researcher leaders indicate the importance of the DAG 1998 evaluation as neutral (mean importance of 3.1). The (follow-up) accreditation of research schools is of minor importance (mean importance 2.8). This view on importance does not depend on organisational setting; the difference in importance of the DAG 1998 evaluation between leaders of university units and leaders of (non-university) institute units is not significant. Not surprisingly, collaboration with national and international research groups is very important for medical and health research groups (mean 7.0). Table 46 shows that, as expected, differences in importance of collaboration with ministries and semi-governmental organisations (F = 10.92; P<0.01), collaboration with the clinic (F = 4.99; χ^2 = 11.59; P<0.05) and collaboration with family doctors (F = 6.57; χ^2 = 11.09; P<0.01) between research groups with a para-clinical, pre-clinical and clinical background are significant. Para-clinical research groups consider collaboration with ministries and semi-governmental organisations as reasonably important (mean 5.8). Pre-clinical and clinical research groups indicate collaboration with ministries and semi-governmental organisations as not important (mean importance 3.2). In addition, para-clinical research groups also find collaboration with general practioners relatively important (mean importance 4.7). This is significantly higher than in pre-clinical research groups, which indicates that collaboration with family doctors is a matter of minor importance (mean 2.2). Pre-clinical and clinical research groups that allocate more time to patient care (see Table 44) find collaboration with the clinic a matter of major concern (mean importance 7.2 and 8.2 respectively). This is significantly higher than for para-clinical research groups (mean importance 6.6).

EXTERNAL CONTROL	PARA-CLINICAL (N = 19)	PRE-CLINICAL (N = 73)	CLINICAL (N = 61)	F VALUE
Time spent on external research activities (6-point scale)	2.3 (0.8)	2.3 (0.6)	2.3 (0.4)	0.68
Time spent on external management activities (6-point scale)	1.4 (1.0)	1.6 (1.2)	1.4 (1.0)	0.69
Importance of externally organised research evaluations (5-point scale)				
DAG evaluation 1998	3.2 (0.6)	3.1 (0.8)	3.0 (0.7)	0.94
Research school evaluations	2.8 (1.1)	3.0 (0.8)	2.8 (1.0)	1.39
Importance of collaboration (10-point scale)				
Research groups	6.9 (1.6)	7.1 (1.7)	7.1 (1.3)	0.15
Ministries and semi-governmental organisations	5.8 (2.4)	3.2 (2.2)	3.2 (2.1)	10.92**
Pharmaceutical industries & firms	3.5 (2.3)	4.7 (2.3)	4.1 (2.3)	1.99
Clinic	6.6 (2.6)	7.2 (2.2)	8.2 (1.9)	4.99**
Family doctor	4.7 (3.1)	2.2 (2.1)	3.4 (3.1)	6.57**

Table 46: Comparison of external control items in para-clinical, pre-clinical and clinical research groups (see 4.3.1 for a description of the measurement scales used)

6.2.4 Research performance

The seven items measuring 'research performance' were compared among para-clinical, pre-clinical and clinical research groups.

Whereas medical and health research leaders of para-clinical groups published an average of 13 SCI papers during the past three years, leaders of pre-clinical and clinical research groups published an average of 25 to 27 papers in SCI journals. This difference between the three groups is not significant due to the large standard deviations. Table 47 shows that most of the medical and health research groups are part of a main research theme that has been externally evaluated in the DAG 1998 as 'very good'. However, it should be noted that only a few para-clinical research groups were involved in the externally organised DAG 1998 evaluation.

The proportion of external funding is high in para-clinical, pre-clinical and clinical research groups; on average, 50% of the personnel and material resources stems from external funding. The number of grants applications submitted to and grants received from the NWO and NWO-Council for Medical and Health Research (MW-NWO) differs (but not significantly) between the three types of medical and health research groups. It seems that pre-clinical research groups submit, on average, more research proposals to NWO (3.1 proposals) and MW-NWO (1.0 proposal) than para-clinical (2.3 NWO proposals, 0.3 MW-NWO proposals) and clinical research groups (1.8 NWO proposals, 0.6 MW-NWO proposals). It is possible that pre-clinical research groups consider NWO and MW-NWO as a more important source of money for receive external research funding than para-clinical and clinical research groups. Finally, the average NWO rating of the research group leaders is high (3.1 measured on 4-point scale). No significant differences among para-clinical, pre-clinical and clinical research groups were found.

RESEARCH PERFORMANCE	PARA-CLINICAL (N = 19)	PRE-CLINICAL (N = 73)	CLINICAL (N = 61)	F VALUE
Research output				
SCI publications (number)	13.2 (10.6)	24.9 (21.8)	27.7 (29.1)	2.49
DAG score (8-point scale)	5.5 (1.6)	6.1 (1.8)	5.9 (1.9)	0.46
Research input				
External funding (percentage)	51.5 (23.3)	51.1 (21.9)	50.5 (25.6)	0.02
Proposals to NWO (number)	2.3 (2.4)	3.1 (4.2)	1.8 (2.1)	2.34
Proposals to MW-NWO (number)	0.3 (0.6)	1.0 (1.5)	0.6 (0.9)	2.96
MW-NWO grants (number)	0.2 (0.4)	0.6 (1.0)	0.3 (0.7)	2.71
NWO rating (4-point scale)	3.3 (1.0)	3.1 (1.2)	3.0 (1.1)	0.31

Table 47: Comparison of research performance in para-clinical, pre-clinical and clinical research groups (see 4.3.1 for a description of the measurement scales used)[70]

6.2.5 Bivariate correlations of variables

In this section the results of three bivariate analyses are presented. First, the correlations between contingency variables and performance variables are reported. Second, the correlations between research management and performance variables are shown. Finally, the correlations between contingency variables and performance variables are given. Only significant correlations are presented in tables. Annexes 5-10 show detailed correlation matrices of the relevant variables.

This section and the remaining ones in this chapter only discuss the results of analyses carried out for pre-clinical and clinical research groups. Unfortunately, it was not sensible to perform bivariate and multivariate analysis of the data of para-clinical groups. Only 13.8% of the response sample (N = 22) could be classified as para-clinical research groups. The size of this subgroup is relatively small, compared to the number of responding pre-clinical (N = 75) and clinical research groups (N = 63). In addition, among the respondents, only a few (5 or 6) para-clinical research groups were involved in the DAG 1998 assessment. Furthermore, it turned out that most of the leaders of para-clinical research groups had submitted hardly any research proposals to NWO and MW-NWO and had received hardly any grants from MW-NWO (see Table 48). For these reasons it was decided not to include para-clinical research groups in the bivariate and multivariate analyses.

[70] Unfortunately, there is no information known about the amount of full-time equivalents (fte) that medical and health groups are spending to conduct medical and health research. Therefore the seven research performance measures cannot be corrected for the total research input. It should be noted that differences with regard to the

6.2.5.1 Bivariate correlations between contingencies and research performance

Table 48 shows significant correlations between contingencies and the research performance of pre-clinical and clinical research groups.

Size

In both pre-clinical and clinical research groups, group size correlates positively with each research performance measure. The size of pre-clinical research groups correlates positively with the number of NWO proposals submitted, the number of MW-NWO proposals submitted and number of MW-NWO grants received, while in clinical research groups size correlates positively with the number of SCI publications and the DAG 1998 score.

Time allocation

The time leaders allocated to research and clinical patient care shows an inverse relationship in pre-clinical research groups. Time allocated to research correlates positively with the DAG 1998 score, the number of MW-NWO grants received and the NWO rating of the research leader. Time allocated to patient care correlates inversely with these three performance measures. Furthermore, time allocated to patient care correlates negatively with the amount of external research funding and the number of NWO research proposals submitted. The amount of time allocated to the supervision of PhD students in pre-clinical research groups is positively associated with the number of SCI publications, the DAG 1998 score, the percentage of external research funding and the NWO rating of the research leader.

Age

The research management experience of pre-clinical research group leaders correlates negatively with their NWO rating. In pre-clinical groups, foreign experiences of the research leader are positively associated with the number of NWO and MW-NWO proposals submitted and the number of MW-NWO grants received. In pre-clinical as well as clinical research groups, the foreign experience of the leader correlates positively with the number of SCI publications. The supervising experience of the medical and health group leader correlates negatively in pre-clinical groups, and positively in clinical research groups, with the number of MW-NWO grants received.

Work setting

Work setting (dummy; code 0= university setting; code 1= non-university institute) of pre-clinical research groups has a positive correlation with the DAG 1998 score and the amount of external

---

amount of fte spend on research between and within para-clinical, pre-clinical and clinical research groups are possible.

research funding. It correlates negatively with the number of SCI publications published during the years 1999-2001.

In sum, it appears that there are both significant positive and negative correlations between contingencies and research performance (Table 48). In pre-clinical research groups significant correlations between contingencies and research performance are more prominent than in clinical research groups.

	PRE-CLINICAL RESEARCH GROUPS	**CLINICAL RESEARCH GROUPS**
SCI publications	+ Time allocated to supervision of PhD students (0.31**) + Foreign experience (0.24*) - Work setting (-0.27*)	+ Size of research group (0.50**) + Time allocated to research (0.42**) + Foreign experience (0.31*)
DAG score	+ Work setting (0.35*) + Time allocated to research (0.67***) + Time allocated to supervision (0.51***) - Time allocated to patient care (–0.63***) - Time allocated to education (–0.31*)	+ Size of research group (0.36*)
External research funding	+ Time allocated to supervision of PhD students (0.27*) - Time allocated to patient care (–0.47***)	+ Supervising experience (0.27*)
NWO proposals submitted	+ Size of research group (0.46***) - Time allocated to patient care (–0.26*)	
MW-NWO proposals submitted	+ Size of research group (0.46**)	
MW-NWO grants received	+ Size of research group (0.38***) + Time allocated to research (0.27*) - Time allocated to patient care (–0.24*) - Research experience (–0.27*) - Supervising experience (-0.27*)	+ Supervising experience (0.29*)
NWO rating	+ Time allocated to research (0.30*) + Time allocated to supervision of PhD students (0.29*) - Time allocated to patient care (–0.33*) - Management experience (–0.31*)	

Table 48: Significant correlations between contingencies and research performance[i]

6.2.5.2 Bivariate correlations between managerial control and research performance

Table 49 shows the significant correlations between the views and judgements of research leaders about research management items (internal and external control) and the research performance of both pre-clinical and clinical research groups. Furthermore, significant correlations between research management activities (internal and external control) undertaken by research leaders and the research performance of both pre-clinical and clinical research groups are shown in this table.

Internal control

Control of personnel resources

Table 49 shows that the extent to which pre-clinical research group leaders control their personnel (i.e. do not experience problems in securing qualified scientific staff) is positively associated with the number of MW-NWO proposals submitted.

Time spent on internal managerial control

In pre-clinical units, the time spent on internal managerial control is positively associated with both the percentage of external research funding received and the NWO rating. In clinical research groups a positive correlation was found between the amount of time allocated to internal managerial control and the number of MW-NWO grants received.

Research commitment

The involvement of pre-clinical research leaders in research activities within their own research groups shows a positive association with the number of SCI publications, the DAG 1998 score and the NWO rating.

Communication

The frequency of internally organised research meetings in pre-clinical research groups correlates positively with number of SCI publications and DAG 1998 score. In clinical research groups, the association between the frequency of internal research meetings and the number of SCI publications is only slightly positive.

Research policy planning

In pre-clinical research groups, discussion meetings about long-term research policy of the group are positively associated with the number of research proposals submitted to the NWO and the NWO rating.

Rewards

In pre-clinical research groups, the extent to which opportunities are offered to develop research skills correlates positively with the DAG 1998 score and the percentage of external research funding obtained, while in clinical research groups research skills correlate positively with the number of SCI publications. In pre-clinical research groups, the extent to which opportunities are offered to work flexible hours is positively associated with the number of NWO and MW-NWO proposals submitted. In contrast, in clinical groups the opportunity to work flexibly is negatively associated with the number of SCI publications. Special commendations given to motivate research staff show, in both pre-clinical and clinical research groups, positive correlations with the number of SCI publications and the amount of external research funding obtained. Furthermore, in pre-clinical groups a positive association is to be found with the number of NWO research proposals submitted. Finally, a positive

correlation was found between the financial rewards offered in clinical groups and the number of SCI publications.

Internally organised research evaluations

The internally organised pre-evaluation of research proposals shows a positive association with the number of SCI publications in pre-clinical research groups. Pre-evaluations of research proposals that are organised on the level of universities and non-university research institutes show a positive correlation with the amount of external funding obtained by pre-clinical research groups. In contrast, in pre-clinical groups the correlation of the pre-evaluation of research proposals internally organised on a faculty level with the number of MW-NWO received grants is negative. The research leaders' opinion that the internal organisation of pre-evaluations of research proposals is effective shows a positive association with the frequency of submission of NWO proposals in clinical research groups. There seems to be a positive association between the internal organisation of research output evaluations (organised on the level of both research groups and the faculty) and the amount of research funding obtained from external funding agencies.

Importance of internally organised research evaluations

The importance of internally organised research evaluations does not correlate significantly with the research performance of pre-clinical and clinical groups.

	PRE-CLINICAL RESEARCH GROUPS	**CLINICAL RESEARCH GROUPS**
SCI publications	+ Special commendations (0.31*) + Research commitment (0.26*) + Research meetings (0.24*) + Pre-evaluation research proposals organised on level of research group (0.24*)	+ Research meetings (0.33*) + Financial bonus system (0.32*) - Flexibility (–0.31*) + Research skills (0.27*) + Special commendations (0.27*)
DAG score	+ Research meetings (0.44**) + Research commitment (0.43**) + Research skills (0.31*)	
External research funding	+ Pre-evaluation research proposals on the level of a university and a non-university research institute (0.41***) + Special commendations (0.35**) + Time spent on internal management activities (0.26*) + Research skills (0.32*)	+ Special commendations (0.42**) + Internal evaluation of research output organised on a faculty level (0.34*) + Internal evaluation of research output organised on the level of research group (0.28*)
NWO proposals submitted	+ Special commendations (0.29*) + Flexibility (0.28*) + Research policy planning (0.25*)	+ Effectiveness of pre-evaluations of research proposals (0.29*)
MW-NWO proposals submitted	+ Flexibility (0.26*) + Resources for control of personnel (0.26*)	
MW-NWO grants received	+ Pre-evaluation research proposals on the level of an faculty/research institute (-0.27*)	+ Time spent on internal management activities (0.27*)
NWO rating	+ Time spent on internal management activities (0.37**) + Research policy planning (0.32**) + Research commitment (0.27*)	+ Internal evaluation of research output on level of faculty/research institute (0.25*)

Table 49: Significant correlations between internal control and research performance

In sum, significant positive correlations are found between internal research management activities and research performance in both pre-clinical and clinical groups. In addition, significant positive correlations are found between the views and judgements of research leaders about internal research management items and the research performance of both pre-clinical and clinical research groups.

External control

Time spent on external research activities

Table 50 shows that the amount of time leaders of pre-clinical groups spend on external research activities, correlates positively with the number of NWO proposals submitted. In clinical groups, a strong positive correlation is found between the time a research leader spends on external research activities and the number of SCI publications.

Time spent on external managerial control

The time pre-clinical research leaders spend on external managerial control shows a positive relationship with the percentage of external research funding obtained and the number of NWO research proposals submitted. In clinical research groups, the time research leaders spend on external management activities shows a positive association with the amount of external research funding obtained and the DAG 1998 score.

Importance of externally organised research evaluations

In pre-clinical research groups the importance of research school assessments correlates positively with the NWO rating ($r = 0.25$*).

Importance of collaboration

In pre-clinical research groups, the interest in collaboration with Ministries and semi-governmental organisations is negatively associated with the DAG 1998 score. In clinical research groups, by contrast, the interest in collaboration with Ministries and semi-governmental organisations correlates positively with the DAG 1998 score. In clinical groups, the interest in collaboration with pharmaceutical industries correlates positively with the DAG 1998 score and the percentage of external funding. In pre-clinical research groups, the interest in collaboration with the clinic is negatively associated with the DAG 1998 score and the amount of external funding. The interest of clinical groups in collaboration with the clinic is negatively associated with the NWO rating. Finally, in pre-clinical research groups, a negative association is shown between the interest in collaboration with family doctors and both the DAG 1998 score and the NWO rating.

In sum, significant correlations are found between the views and judgements of research leaders about external collaboration and the research performance of both pre-clinical and clinical research groups.

It seems that the importance attached to collaboration with the clinic, general practitioners, Ministries and semi-governmental organisations correlate negatively with the research performance of pre-clinical groups. In both pre-clinical and clinical research groups significant positive correlations are found between external research management activities and research performance.

	PRE-CLINICAL RESEARCH GROUPS	**CLINICAL RESEARCH GROUPS**
SCI publications		+ External research activities (0.47**)
DAG score	- Importance of collaborations with clinic (–0.35*) - Importance of collaborations with Ministries and semi-governmental organisations (–0.31*) - Importance of collaborations with family doctor (–0.30*)	+ Importance of collaborations with pharmaceutical industries & firms (0.39*) + Time for external managerial control (0.36*) + Importance of collaborations with Ministries and semi-governmental organisations (0.33*)
External research funding	+ Time spent on external managerial control (0.37**) - Importance of collaboration with clinic (–0.27*)	+ Time spent on external managerial control (0.40**) + Importance of collaboration with pharmaceutical industries/firms (0.29*)
NWO proposals submitted	+ Time spent on external managerial control (0.32**) + Time spent on external research activities (0.28*)	
MW-NWO proposals submitted		
MW-NWO grants received		
NWO rating	- Importance of collaboration with family doctor (-0.26*) + Importance of research school accreditation (0.25*)	- Importance of collaboration with clinic (–0.26*)

Table 50: Significant correlations between external control and research performance

6.2.5.3 Bivariate correlations between contingencies and managerial control

Contingencies, such as the size of the research group, the time that can be spent on research and the experience of the research leader, may influence the choice and the extent to which research management activities are organised and introduced in medical and health groups. Also, contingencies may influence the views and judgements of research leaders about both internal and external research management items. Table 51 presents significant correlations between the contingencies of both pre-clinical and clinical research groups and research management items.

Size

The number of staff employed in health research groups is positively correlated with the time the research leader spends on external management activities and with the extent to which the research leader uses financial bonus systems to motivate employees. Furthermore, the size of clinical research groups shows a positive correlation with the time research leaders spend on external research activities and with the extent to which research leaders offer special commendations. In pre-clinical research groups a positive correlation is found between group size and the organisation of research policy meetings.

Time allocation

In pre-clinical research groups, time allocation shows an inverse relationship with the importance of collaboration with family doctors, clinics and pharmaceutical industries/firms. If more time is allocated to research and the supervision of PhD students, the importance of collaboration between pre-clinical research groups and family doctors, clinics and pharmaceutical industries/firms is lower. If the emphasis is put on patient care, the importance of collaboration with family doctors, clinics and the pharmaceutical industry/firms is higher. Also, an inverse relationship in pre-clinical groups is found between time allocation (to research and patient care) and research commitment.

Furthermore, the time pre-clinical research leaders allocate to research and supervision shows a positive correlation with the number of research meetings. In pre-clinical research, the time allocated to patient care shows negative correlations with the amount of time spent on both internal and external management activities and the extent to which opportunities are offered to staff members to develop their research skills. Negative correlations are also found in pre-clinical groups between the time allocated to research and the level of direct communication and between the amount of time allocated to the supervision of PhD students and the importance attached to internal research evaluations.

In clinical research groups, time allocated to research and supervision is positively associated with the amount of time spent on internal management activities, the amount of time spent on external research activities and research commitment. In clinical research groups, in addition, time allocated to research shows a positive correlation with the research leaders' opinion about the effectiveness of internal pre-evaluation of research proposals. In clinical groups, time allocated to research correlates negatively, and time allocated to patient care positively, to the importance attached to collaboration with the clinic. Moreover, time allocated to patient care is negatively associated with the time spent on internal management activities and the extent to which staff are offered the opportunity to work flexibly.

Age

The research experience of pre-clinical leaders correlates negatively with the extent to which opportunities are offered to staff to work flexibly and with the importance attached to collaboration with pharmaceutical industries and firms. The management experience of pre-clinical leaders is positively associated with the importance attached to collaboration with family doctors and negatively with the number of research meetings organised in the responding groups. In clinical research groups, the management experience of research leaders seems to have a positive association with the personnel resource control and with the organisation of discussions of internal research policies. Table 51 shows that, in pre-clinical research groups, the foreign experience of leaders is positively associated with research commitment. In clinical research groups, the foreign experience of research leaders correlates positively with:

– the time spent on external research activities

– the extent to which opportunities are offered to develop research skills

– the extent to which special commendations are given to staff members, and
– the importance attached to collaboration with pharmaceutical industries and firms.
The experience of pre-clinical research leaders in supervising the research group is negatively associated with the extent to which financial bonus systems are used to stimulate staff. Also, a negative association is found with the importance attached to collaboration with pharmaceutical industries/firms, governmental agencies and family doctors. The experience of clinical leaders in supervising their research group show a positive correlation with the use of an internal evaluation of research output on the level of the group.

Work setting

The work setting is only correlated with research management items in pre-clinical research groups. It correlates positively with the organisation of pre-evaluation of research proposals on an organisational level, as well as with the opinion about the effectiveness of those evaluations as shown in Table 51. The work setting is also positively associated with the research commitment of pre-clinical leaders. Negative correlations are found between the work setting and two management activities (meetings about research policy and using of financial bonuses) and two managerial opinions about research collaboration.

In sum, it is shown that contingencies have significant associations with views and judgements of research leaders about both internal and external research management items. Contingencies also have significant associations with the use of internal and external research management activities. In pre-clinical research groups, positive correlations are found between contingencies and research management items except for the contingencies and research management items related to patient care. In clinical research groups, contingencies show mainly positive correlations with research management (both views/judgements and activities). In both types of research groups the amount of time allocated to patient care shows a negative correlation with research management views/judgements and activities.

	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
Staff members	+ Financial bonus system (0.34**) + Research policy planning (0.27*) + Time spent on external managerial control (0.25*)	+ Special commendations (0.37**) + Time spent on external managerial control (0.32*) + Time spent on external research activities (0.29*) + Financial bonus system (0.28*)
Work Setting	+ Pre-evaluation of research proposals on university/non-university institute level (0.39*) - Financial bonus system (-0.29*) - Importance of collaboration with family doctor (-0.29*) + Research commitment (0.28*) - Research policy planning (-0.28*) - Importance of collaboration with Ministries and semi-governmental organisations (-0.27*) + Effectiveness of pre-evaluations (0.27*)	
Time allocated to research	+ Research commitment (0.45***) - Direct communication (-0.40**) + Research meetings (0.29*) - Importance of collaboration with family doctor (-0.26*) - Importance of collaboration with clinic (-0.25*)	+ Time spent on internal managerial control (0.35**) + Time spent on external research activities (0.33*) + Research commitment (0.29*) + Effectiveness pre-evaluation proposals (0.26*) - Importance of collaboration with clinic (-0.26*)
Time allocated to education	+ Importance of collaboration with family doctor (0.31*) + Financial bonus system (0.24*) - Importance of collaboration with clinic (-0.25*)	+ Internal evaluation research output on faculty level (0.35*)
Time allocated to patient care	+ Importance of collaboration with Ministries and semi-governmental organisations (0.41***) - Time spent on internal managerial control (-0.37**) + Importance of collaboration with clinic (0.34**) + Importance of collaboration with family doctor (0.34**) - Research commitment (-0.33*) - Time spent on external managerial control (-0.29*) - Development of skills (-0.29*)	+ Importance of collaboration with clinic (0.44**) - Flexibility (-0.32*) - Time spent on internal managerial control (-0.35**)
Time allocated to supervision	+ Research commitment (0.47***) + Time spent on internal managerial control (0.43***) - Importance of collaboration with Ministries and semi-governmental organisations (-0.31*) + Research meetings (0.27*) - Importance of collaboration with family doctor (–0.27*) - Importance of internal research evaluations (–0.25*)	+ Time spent on internal managerial control (0.40**) + Research commitment (0.37**) + Time spent on external research activities (0.27*)
Research experience	- Importance of collaboration with pharmaceutical industries and firms (-0.26*) - Flexibility (-0.24*)	
Management experience	+ Importance of collaboration with family doctor (0.42**) - Research meetings (-0.29*)	+ Resources control: personnel (0.43*) + Special commendations (0.32*)
Foreign experience	- Importance of collaboration with clinic (-0.31*) + Development of skills (0.28*) + Control of personal resources (0.26*) + Special commendations (0.24*)	+ Importance of collaboration with clinic (0.37**) + Importance of collaboration with research groups (0.28*) + Time spent on external research activities (0.26*)
Supervising experience	+ Importance of collaboration with pharmaceutical industries & firms (–0.45***) - Importance of collaboration with Ministries and semi-governmental organisations (–0.35**) - Financial bonus system (–0.33**) - Importance of collaboration with family doctor (–0.25*)	+ Internal evaluation of research output on the level of group/department (0.39**)

Table 51: Significant correlations between contingencies and research management items

6.2.6 Multivariate associations

To explore the causal relationships between contingencies, research management and research performance in pre-clinical and clinical research groups (see Figure 11), stepwise multiple regression analysis was carried out.

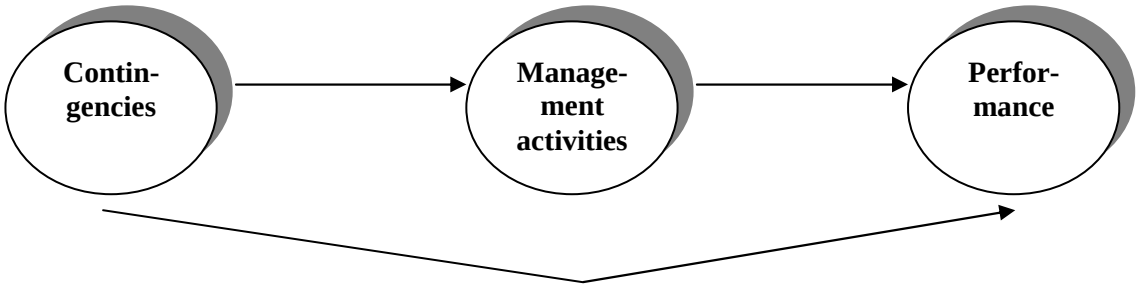

Figure 11: The internal organisation of research groups

A stepwise approach was chosen because of the exploratory character of this research question. The set of independent variables was not closely specified in the regression model. Furthermore, the set of independent variables was not essentially used in a confirmatory approach. A stepwise approach allows the researcher to examine the contribution of each independent variable to the regression model (Hair *et al*., 1998). Each variable was considered for inclusion prior to developing the equation. The independent variable (contingencies and research management items) with the greatest contribution was added first. Independent variables were then selected for inclusion based on their incremental contribution to the variables already present in the equation (Hair *et al*., 1998). The independent variables for which incremental *F* was not significant even at the 5% level were excluded. A series of stepwise regression (forward addition) analyses was carried out to determine the sets of predictors separately for each performance measure. The results are presented in Tables 53-59. R^2 (adj) indicates the variance explained by the predictors.

Validity

Before stepwise regressions were conducted, the construct validity of outcome variables was assessed by using factor analysis (Principal Component Analysis, PCA). Table 52 shows that three factors, with a value greater than 1.00, were extracted. These three factors explain 74.45% of total variance. The PCA analysis revealed one factor of three items – the number of NWO proposals submitted, the number of MW-NWO proposals submitted and the number of MW-NWO grants obtained – reflecting the NWO activity of research groups (component 1). These three items can be reduced to one dimension. A one-dimensional index, 'NWO activity', was constructed and used in multivariate

stepwise regression analysis. There are two items that load on more than one component. The variable 'DAG 1998 score' has a low loading on both component 2 and component 3. The variable 'NWO rating' has a medium loading on both component 1 and component 2. In total 5 different variables were used in the stepwise regression analysis to measure research performance.

	COMPONENT 1 FACTOR LOADINGS	**COMPONENT 2 FACTOR LOADINGS**	**COMPONENT 3 FACTOR LOADINGS**
Number of SCI publications	0.09	0.08	0.93
DAG 1998 score	0.20	0.44	0.47
External research funding	0.02	0.88	0.03
Number of NWO proposals submitted	0.78	0.38	0.14
Number of MW-NWO proposals submitted	0.94	0.12	0.07
Number of MW-NWO grants obtained	0.91	0.03	0.15
NWO rating	0.45	0.62	0.09
Explained variance	**36.59**	**21.43**	**16.43**

Table 52: Factor loadings and variance explained after varimax rotation: results of a principal component analysis of seven items for research performance of all research groups

Number of SCI publications

Stepwise multiple regression analysis shows three predictors explaining about 27% of the variance in number of SCI publications (see Table 53) within pre-clinical research groups. The amount of time a research leader allocates to supervision of PhD students is the most important predictor ($\beta = 0.36$; $P<0.001$). Also special commendations given to research staff ($\beta = 0.32$; $P<0.01$) predict the number of SCI publications. Interestingly, the work setting[71] of research groups has a negative relationship with the number of SCI publications ($\beta = -0.38$; $P<0.001$). This means that working in a non-university institute has a negative relationship with the number of SCI publications that are published in these groups. This result may be explained by existing differences in publication strategies (e.g. Moed, 2000) between research groups. Open interviews that are conducted –in the earlier stages of this study- with experienced pre-clinical research leaders of non-university institutes showed that a number of Dutch medical and health non-university research institutes are stimulating researchers to focus on quality. They are encouraged to publish in highly renowned SCI journals within their medical subfield or to publish in interdisciplinary prestigious journals such as Nature, Science and Cell. As a

[71] Work or organisational setting (university versus institute) is used in this study as a contingency variable. In 1995 Omta demonstrated that institutes judge elements of system control (constituted as an element of managerial control) e.g. effectiveness of personnel policy, adequacy of resources and administrative control more positively compared to universities.

consequence, the number of SCI publications of these research groups (quantity) may be relatively low compared to research groups that are not managed in this way.

The step-wise regression reveals that 65% of the variance in the number of SCI publications within clinical research groups can be explained by seven predictors (Table 53). The three most important predictors are contingency variables: (1) the size of the research group (β = 0.44; P<0.001), (2) the amount of time allocated to research activities (β = 0.46; P<0.01) and (3) the amount of time allocated to patient care (β = 0.38; P<0.001). In addition to contingencies, four research management variables account for a fair amount of the variance in the variable in question. The first is the opinion of the clinical research group leaders about the importance of internal research assessments (β = –0.24; P<0.01). The second is the amount of time clinical research leaders spend on external research activities (β = 0.25; P<0.01). The third is the control of personnel resources (β = 0.18; P<0.05). The fourth research management variable is the frequency of internal research meetings (β = 0.17; P<0.05).

RANKING OF PREDICTORS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
1st rank	Time allocated to supervision (β = 0.36***)	Research staff (β = 0.44***)
2nd rank	Work setting (β = -0.38*)	Time allocated to research (β = 0.46**)
3rd rank	Rewards: special commendations (β = 0.32*)	Time allocated to patient care (β = 0.38***)
4th rank	–	Importance internal research evaluations (β = –0.24**)
5th rank	–	Time spent on external research activities (β = 0.25**)
6th rank	–	Control of personnel resources (β = 0.178*)
7th rank	–	Research meetings (β = 0.20*)
R^2 (adj)	0.27	0.65
F ratio	9.57***	17.12***

Table 53: Step-wise regression of contingency and research management variables on the number of SCI publications. $^{*}P$<0.05; $^{**}P$<0.01; $^{***}P$<0.001

DAG 1998 score

In pre-clinical groups, time allocated by a research leader to research and patient care are the most important predictors for the DAG 1998 score (see Table 54). Time allocated to research has a positive relationship with the DAG 1998 score. However, time allocated to patient care has a negative relationship with the DAG 1998 score. Furthermore, three research management variables have linear relationships with the DAG 1998 score. The number of research meetings shows a positive relationship (β = 0.29; P<0.01) with the DAG 1998 score: more research meetings will result in a higher DAG 1998 score. Research output evaluations internally organised by universities and non-university research institutes also have a positive relationship with the DAG 1998 score (β = 0.24; P<0.01). Interestingly, research output evaluations organised by the research group itself show a negative relationship with the DAG 1998 score (β = –0.21*; P<0.05). The set of predictors explains

about 54% of the variance in the DAG 1998 score. In clinical groups, the set of predictors was comprised of indices reflecting the time that is spent on external management activities and the interest that is taken in collaboration with governmental agencies. These predictors together explain 22% of the variance in the DAG 1998 score.

RANKING OF PREDICTORS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
1st rank	Time allocated to research (β = 0.28**)	Time spent on external management activities (β = 0.39**)
2nd rank	Time allocated to patient care (β = –0.38***)	Importance collaboration with pharmaceutical industries (β = 0.34***)
3rd rank	Research meetings (β = 0.29**)	-
4th rank	Research output evaluations internally organised by universities and non-university research institutes (β = 0.24**)	–
5th rank	Research output evaluations internally organised by research groups (β = –0.21*)	–
R^2	0.54	0.22
F ratio	15.31***	8.07***

Table 54: Step-wise regression of contingency and research management variables on the DAG 1998 score. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

External research funding

Table 55 shows that if a pre-clinical research group leader allocates more time to patient care this will decrease the amount of external research funding obtained (β = –0.45; $P<0.001$). Moreover, the use of pre-evaluations of research proposals -before they are submitted to external funding agencies- that are internally organised by universities and non-university research institutes has a positive relationship with the percentage of external research funding obtained (β = 0.37; $P<0.001$). Also the extent to which special commendations are given to research staff is positively related to the percentage of external research funding. These three predictors together explain 42% of the variance in percentage of external research funding obtained. The set of external research funding predictors in clinical research groups, as presented in Table 55 by step-wise regression, comprises four management activities. Together they explain 33% of the variance.

1. Amount of time clinical research leaders allocate to external management activities (β = 0.28; $P<0.05$);
2. Extent to which special commendations are used to stimulate staff members (β = 0.31; $P<0.01$);
3. Evaluation of research output internally organised by medical faculties (β = 0.35; $P<0.01$);
4. Interest in collaboration with family doctors (β = 0.24; $P<0.05$). One possible explanation of this last result might be that family doctors, because they have access to their own extensive patient database, are easily able to select and find patients with particular diseases or syndromes. Collaboration between researchers and family doctors may contribute to solving sample-size problems

in study designs. Lastly, initiation and maintenance of collaboration with family doctors can be interpreted as an external research management activity. However, no significant correlation between time allocated to external management activities and interest in collaboration with family doctors has been found.

RANKING OF PREDICTORS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
1st rank	Time allocated to patient care (β = –0.45***)	Time spent on external management activities (β = 0.28*)
2nd rank	Pre-evaluations of research proposals internally organised by universities and non-university research institutes (β = 0.37***)	Special commendations (β = 0.31**)
3rd rank	Special commendations (β = 0.25**)	Pre-evaluations of research proposals internally organised by medical faculties (β = 0.35**)
4th rank	–	Importance attached to collaboration with family doctors (β = 0.24*)
R^2	0.42	0.33
F ratio	18.14***	8.43***

Table 55: Step-wise regression of contingency and research management variables on the percentage of external research funding obtained. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

NWO activity

Table 56 shows that the set of predictors of Dutch research-council activity in pre-clinical groups consists of indices of contingencies and internal management activities. Four predictors explain 35% of the total variance. As shown in Table 56, the size of the pre-clinical research groups contributed most to the explained variance in Dutch research-council activity of research groups (β = 0.29; $P<0.001$). Other positive significant linear relations in order of importance are the time research leaders spend on external research activities (β = 0.23; $P<0.05$) and the extent to which flexibility in both working hours and work place (e.g. the possibility of working at home) is offered and used by leaders as a reward to motivate research staff (β = 0.20; $P<0.05$). Not surprisingly, the amount of time research leaders allocate to patient care (β = –0.19; $P<0.05$) shows a negative relationship with Dutch research-council activity.

RANKING OF PREDICTORS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
1st rank	Research staff ($\beta = 0.29$***)	Effectiveness of internally organised pre-evaluation of research proposals ($\beta = 0.27$*)
2nd rank	Time allocated to patient care ($\beta = -0.28$*)	
3rd rank	Time spent on external research activities ($\beta = 0.23$*)	
4th rank	Rewards: flexibility ($\beta = 0.20$*)	
R^2	0.35	0.06
F ratio	10.52***	4.69*

Table 56: Step-wise regression of contingency and research management variables on NWO activity. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

By using univariate analysis of variance it can be shown (Table 57) that pre-clinical research groups that organise pre-evaluations of research proposals and do not participate in pre-evaluations of proposals organised by medical faculties, submit more proposals to NWO and MW-NWO and also obtain more MW-NWO grants (mean Dutch research-council activity 3.20) than pre-clinical research groups not involved in pre-evaluations of proposals at all (Dutch research-council activity 1.11). Interestingly, pre-clinical research groups that not only organise pre-evaluations of proposals but also participate in pre-evaluations organised on a faculty level show almost the same Dutch research-council activity (mean 1.18) as pre-clinical research groups not involved in pre-evaluations of proposals (mean 1.11). Apparently, involvement in pre-evaluations of proposals at both levels of organisation causes an additional barrier to undertaking Dutch research-council activity. For example pre-evaluations can result in inconsistency in evaluation scores, comments and suggestions to improve research proposals. Furthermore, pre-evaluation processes take time. To summarise, researchers can be discouraged from submitting research proposals. This process is further amplified given the fact that the Dutch research-council success percentage is already relatively low.

PRE-EVALUATION OF RESEARCH PROPOSALS		NWO ACTIVITY (MEAN)	N
Research group level	Medical faculty level		
0 (No)	0 (No)	1.11	12
1 (Yes)	0 (No)	3.20	15
1 (yes)	1 (Yes)	1.18	41
0 (No)	1 (Yes)	–	0

Table 57: Univariate analysis of variance with NWO activity as a dependent variable

Table 56 indicates that there is only one variable that is related to Dutch research-council activity in clinical research groups: research leaders who think that internally organised pre-evaluations of research proposals are effective submit more proposals to both the NWO and the MW-NWO and obtain more MW-NWO research grants ($\beta = 0.27$; $P<0.05$). This predictor explains only 6% of the variance in Dutch research-council activity. Of more interest is the observation that, in the period

1999–2001, no less than 34% of responding clinical research groups did not submit research proposals to either the NWO or the MW-NWO. In order to investigate further the relation between contingencies, research management and Dutch research council activity in clinical research groups, a dummy variable was constructed. A dummy is used because the variable is not normally distributed (code 0= no Dutch research-council activity; code 1= any Dutch research-council activity). Table 58 indicates that there are three management activities that are positively related to Dutch research-council activity in clinical research groups: (1) stimulation of research staff by offering special commendations (β = 0.33; P<0.01); (2) participation in research output evaluations organised by medical faculties (β = 0.28; P<0.05) and (3) organisation of internal research policy meetings (β = 0.26; P<0.05). The importance of collaboration with the clinic has a negative relationship with Dutch research council activity (β = -0.30*; P<0.05). These factors together explain 26% of the variance in Dutch research-council activity. The mission of the Dutch research council in the period 1999-2001 to stimulate mainly basic medical and health research can explain the negative relationship between collaboration with clinic practice and Dutch research-council activity. Also the criteria –mainly scientific quality- that were used in the selection process favoured the stimulation of pre-clinical research.

RANKING OF PREDICTORS	**CLINICAL RESEARCH GROUPS**
1st rank	Rewards: special commendations (β = 0.33**)
2nd rank	Research output evaluations internally organised by medical faculties (β = 0.28*)
3rd rank	Importance of collaboration with clinics (β = –0.30*)
4th rank	Research policy planning (β = 0.26*)
R^2	0.26
F ratio	6.19***

Table 58: Step-wise regression of contingency and research management variables on NWO activity in clinical research groups. $^{*}P$<0.05; $^{**}P$<0.01; $^{***}P$<0.001

NWO rating of research group leaders

Table 59 indicates that four research management variables are related to the NWO rating of pre-clinical research leaders: (1) the amount of time leaders spend on internal management activities (β = 0.32^{**}), (2) the organisation of internal research policy meetings (β = 0.50^{***}), (3) participation in research output evaluations internally organised on the level of the medical faculties (β = $–0.29^{*}$) and (4) direct communication between research group members (β = -0.23^{*}). Furthermore, the management experience of the pre-clinical leader has a negative relationship with the NWO rating (β = $–0.24^{*}$). These four research management variables and one contingency factor together explain 42% of the variance in the NWO rating. In clinical research groups, four research management variables have a relationship with the NWO rating of the leaders. The internal organisation of policy as well as

research meetings has a positive relationship with the NWO rating ($\beta = 0.27^{*}$ and $\beta = 0.25^{*}$ respectively). Furthermore, a positive relationship has been found with participation in research output evaluations organised by medical faculties ($\beta = 0.27^{*}$). The most important predictor of the NWO rating of pre-clinical research group leaders seems to be the importance attached to collaboration with the clinic. The importance given to collaboration with the clinic has a negative relationship with the NWO rating of the research leader. The four factors together explain 26% of the variance in the NWO rating of clinical research group leaders (see Table 59 for an overview).

RANKING OF PREDICTORS	PRE-CLINICAL RESEARCH GROUPS	CLINICAL RESEARCH GROUPS
1st rank	Time spent on internal management activities ($\beta = 0.32^{**}$)	Importance of collaboration with clinic ($\beta = -0.32^{**}$)
2nd rank	Research policy planning ($\beta = 0.50^{***}$)	Research policy planning ($\beta = 0.27^{*}$)
3rd rank	Research output evaluation internally organised by medical faculties ($\beta = -0.29^{*}$)	Research output evaluation internally organised by medical faculties ($\beta = 0.27^{*}$)
4th rank	Direct communication ($\beta = -0.23^{*}$)	Research meetings ($\beta = 0.25^{*}$)
5th rank	Research commitment ($\beta = 0.20^{*}$)	-
R^2	0.42	0.26
F ratio	9.64***	5.03***

Table 59: Step-wise regression of contingency and research management variables on the NWO rating of research group leaders. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

6.3 Concluding remarks

We have analysed the relationship between experiences with and views of managerial control and the performance of Dutch medical research groups. Our empirical study shows that there is one internal research management activity which has a positive relationship with medical research performance in general. Offering honorable mentions to medical (both pre-clinical and clinical) research staff members, including non-financial prizes, in order to motivate them is positively correlated with the number of SCI publications, the percentage of externally obtained research funding and the NWO activity of research groups. In the regression analyses with contingencies and managerial control variables we found a positive linear relationship between special commendations and two medical research performance measures. It was found that the disciplinary setting, and classification into pre-clinical and clinical research, affect the relationships between managerial control and performance. Interesting differences were found between pre-clinical and clinical research groups. Section 7.3 presents and discusses these differences in more detail (see Table 62).

Patient care and performance

As could be expected, the time allocated to research and the time allocated to patient care show an inverse relationship. Research leaders of pre-clinical groups spend significantly more time on research than research leaders of clinical groups. However, this extra time spent is not translated into better research performance. These results are in line with the work of Omta (1995) who found, in a study of medical scientific staff, that although scientific staff in pre-clinical units can spend twice as much time on research as staff in clinical units, this does not result in higher research quality or user effectiveness. In addition, we found negative linear relationships between the time pre-clinical research leaders spent on patient care and various research performance measures: the amount of research funding, NWO activity and the DAG 1998 score. Interestingly, in clinical research groups time that was spent on patient care had a positive linear relationship with the number of SCI publications[72]. An explanation for this unexpected finding is that working in clinical practice enlarges the possibilities of generating interesting and innovative research questions. In this respect the recent bibliometric study of research conducted at Dutch university medical centres (2004) should be mentioned. It shows that the number of publications of UMCs in clinical journals is relatively high. For instance the impact on general and internal medicine is quite impressive; the citation score is almost three times the international average.

Reflection on research performance indicators used

We experienced problems by using the DAG 1998 score as a research-group performance measure. First, research has been evaluated within the DAG 1998 procedure on the level of research themes. Research leaders were asked in the survey study to write down the overall DAG 1998 quality score of the research theme to which their groups belonged. It should be noted that a research theme consists of several (parts of) research groups. Therefore, (small) differences in DAG scores among research groups within a theme are possible, but not known. Second, factor analysis as presented in Table 54, showed that for the DAG 1998 there were only medium loads (0.44 and 0.47) on components 2 and 3 and a low load (0.20) on component 1. This indicates that more than one component explains the variance of the variable, DAG 1998. Finally, the interview part of this thesis showed that both research leaders and research administrators reported negative views and experiences regarding the DAG 1998 procedure. Consequently, the impact and implementation of DAG 1998 outcomes and recommendation were relatively low. In conclusion, the DAG 1998 score is less meaningful than generally assumed.

[72] In search for an explanation for this result we studied the combined effect of two factors (interaction effect). This means that the effect of one variable has to be shaded by the other variable. The positive effect of time allocated to patient care on the number of SCI publications might be valid for leaders who also spent some time on, for example, (internal/external) research activities. For research leaders who spent no time or too much time (in relation to the time allocated to patient care) on research activities, the effect would be absent or even negative. However, no significant effects of interactions were found.

In conclusion, five objective measures are used in this study to indicate research group performance. Although the technical and methodological problems involved with an indicator, for gathering and handling data and its relative use as a measure of research performance have been individually discussed and criticised by scholars, the simultaneous use of various measures as applied in this chapter provides a good assessment of research performance in medical and health groups.

CHAPTER 7: CONCLUSIONS AND DISCUSSION

This thesis studied how Dutch medical and health research groups are affected by both internal and external evaluations and how (and in what way) they react to them. In addition, the study investigated whether research evaluations and other managerial control elements influence research- group performance.

Although a number of scholars have studied the externally organised evaluation processes in the Netherlands, the effects and implications of these evaluations are hardly ever studied. In addition to external evaluations, we also reported information about internally organised research evaluations and their implications. They are interpreted as management activities. This approach is innovative; until now relatively little was known about internal research evaluations. In addition, empirical studies that described the relationship between managerial control and research performance omitted research evaluations from the scope of research. In this thesis research management, known contingency items, and research performance were studied together in one study. This is rarely done. Moreover, while empirical studies -which examined research performance- concentrated on only one or a few research management activities, managerial views, contingencies or personal characteristics of the scientist, this thesis was used to assess the relevance of the various factors known to influence research performance.

This thesis used qualitative (interviews) and quantitative (questionnaire) techniques, as described in detail in chapter 3. It was organised in such a way that the first stages led to further definitions of the research area by finding out what was known about the topic and then developing more detailed questions about the topic for the main part of the empirical work. Both interview protocol and questionnaire were designed (including pre-testing) by obtaining in-depth information through interviews. Also literature was reviewed (see chapter 2). The response rate for the empirical management study was good. Together, this may inspire more confidence in the general validity of the findings.

In this final chapter the general research questions, which were formulated in chapter 1 and presented in the empirical part of this thesis (chapter 4, 5 and 6), will be answered and discussed. The chapter ends with some implications for further research on the management of academic research.

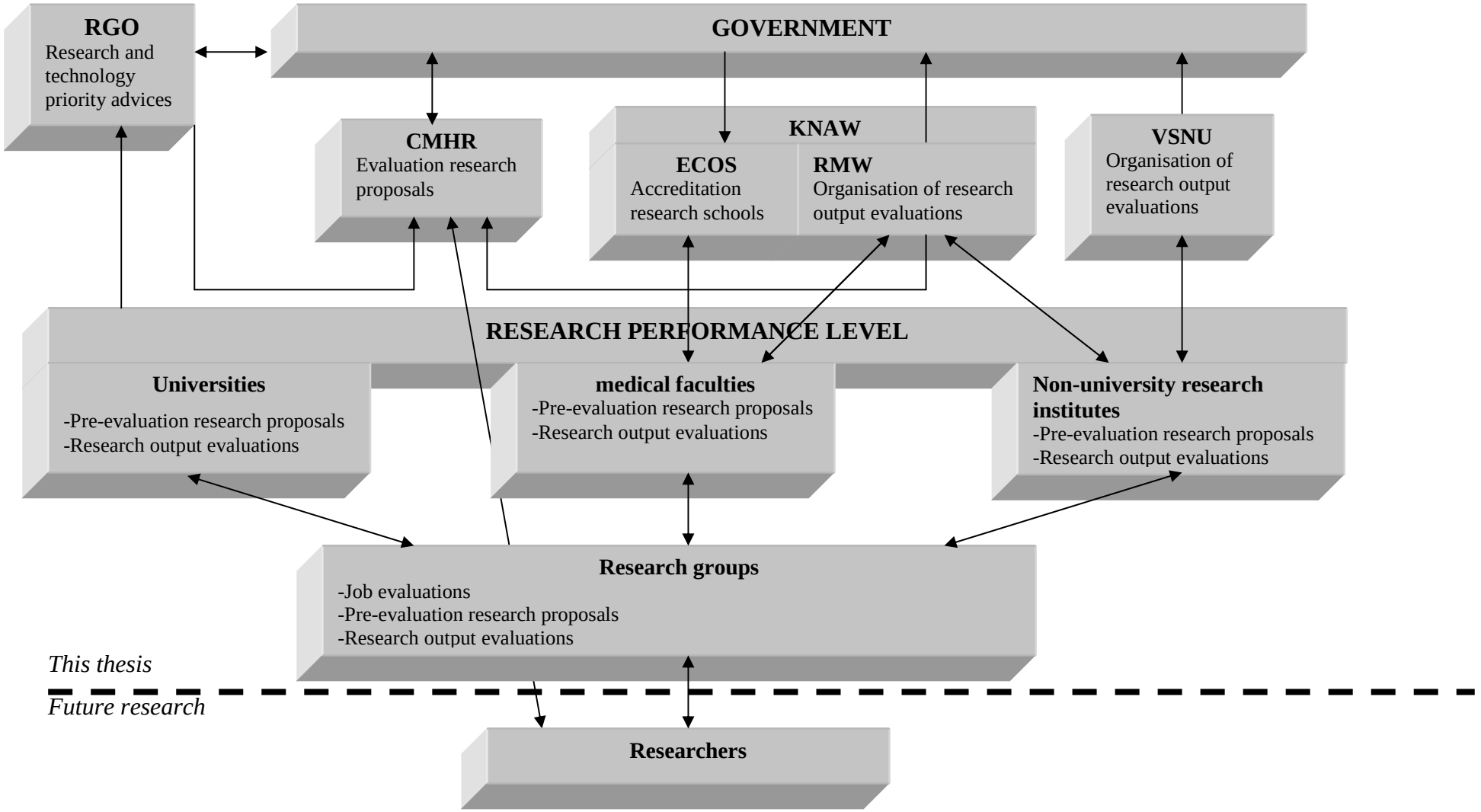

Figure 12: Summary of medical and health research evaluations operating during the time when empirical data for this thesis was collected (2000-2002).

7.1 Medical and health research evaluations in the Netherlands

Chapter 4 showed that medical and health research groups in the Netherlands have to deal with evaluation practices organised by various intermediary organisations in all phases of the research process. Furthermore, medical and health research groups in the Netherlands are subjected to internal research evaluations, as is shown in chapter 5. Different entities such as universities, non-university institutes and medical faculties organise such internal research evaluations. Also the majority of the research leaders interviewed conduct their own evaluations to control and stimulate their research input and research output. Research evaluations studied in this thesis have been summarised in Figure 12. This figure is based on Figure 6 (chapter 4, page 103) and Figure 9 (chapter 5, page 125). The arrows in this figure are described and explained in section 4.5 and section 5.3 of this thesis. This thesis reveals that some of the external research evaluation practices organised by intermediaries are copied and used in internal research evaluations by entities acting on the performance level. The organising bodies operating in the Dutch medical and health systems are subject to various changes in the way they interact, which in turn constantly influences the way evaluations are made. For example, the DAG evaluation no longer exists in the form reported.

We have seen that in all the research evaluation practices encountered in this study there is a traditional evaluation of scientific quality based on peer review. Peer review is a collective noun of different evaluation practices that uses peers. Chapters 4 and 5 showed that peers are asked in ex-ante, ex-post and incidental evaluations (such as RGO advices) to give comments and recommendations to improve research. If research proposals are pre-evaluated, advice about the possible external success percentages is also given.

The protocol of evaluation practices differ, for example the use of peer review committees or individual peers. Questions which organizers have to deal with are for instance: how to select peers; how to compare the pro's and cons of peer independence and expertise; who sets priorities and makes selections and takes final decisions and how is all this being performed? The selection of (anonymous) peers is one of the most important steps in the evaluation process. A peer must be competent to carry out the evaluation. In internal research evaluations direct colleagues or researchers who are employed within the same faculty or non-university research institutes indicate research performance. In external evaluations, peers must also be completely independent from the evaluation object involved. Consequently, only external peers are used. In internal evaluations external peers are rarely involved. In the evaluation of scientific quality of research peers are always scientists.

Additionally, the question of societal quality, impact and relevance is becoming increasingly important, especially in externally organised research evaluations. However, Figure 13 shows that ZonMw is as yet the only intermediary that let peers indirectly evaluate the societal aspects of research. These proposals are submitted to societally relevant research programmes. Non-university

research institutes and universities that have been studied for this thesis do not use societal quality as a performance indicator (see Figure 13). Only one research institute, which has a clear mission to produce 'useful' knowledge and technology, deals with both scientific and societal aspects of medical and health research in their self-evaluations. Methodologies and indicators to measure the societal aspects of medical and health research are now being developed and tested in the Netherlands. Most of these experimental methodologies are quantitative evaluation tools that can be used in qualitative peer practices such as self-evaluations and external audits. This is not solely a Dutch concern; it is also being addressed in other countries. Lyall et al. (2004) for instance recently reported about the assessment methodology of end-user relevance in evaluating the output of biological (and agricultural) research organisations in the UK. Furthermore, societal aspects of research cannot be grafted onto the traditionally based peer review process. Because research activities are not only geared towards scientific colleagues, but also to other stakeholders, traditional peer review processes can be extended by including stakeholders, other than scientists, in the evaluation process of science. In that way, the Medical Committee of the KNAW has proposed the consultation of patient organisations, health-care institutions, policy-makers and research clients. The findings of the shift to extended peer review are in line with discussions in other Dutch studies, as for example the inclusion of lay jury[73] in the evaluation process of research proposals organised by the Dutch Technology Foundation (van der Meulen and Rip, 2000). Finally, it may be possible that by consultating patient organisations, health-care institutions, policy makers and research clients as suggested within the new national research evaluation system (see Figure 13 last column), the societal impact of research will become an increasingly important indicator in both internally and externally organised Dutch medical and health research evaluations.

7.2 Implications of externally organised research evaluations

In the medical and health sector, externally research evaluations are organised to fund research (ZonMw), to mark performance (ECOS), to inform about research (RGO, VSNU, RMW) and to direct research (RGO and ZonMw). The impact of externally organised evaluation procedures and outcomes on Dutch medical and health research differs. As shown in chapter 4, ZonMw evaluation practices and evaluation outcomes have a high impact. A few features of the evaluation procedures are borrowed and used in internally organised *ex-ante* evaluations. Also the winning MW-NWO grants are highly esteemed on the research performance level. In addition, ZonMw evaluation outcomes are used as a research performance indicator in internal resource (re)allocation processes. As reported in chapter 4, interviews about the accreditation of research schools showed that their impact has been reduced. The

[73] In this case, lay in the sense that they need not have any knowledge about the research domain of the research

strong increase in the number of research schools, which have obtained accreditation, seems to be the most important reason. Therefore, an accreditation is now no longer an indicator of performance. The interview and survey study as presented in chapter 4 has revealed that the DAG 1998 evaluation had little impact.

Three different problems were identified within the evaluation procedure, which limited the use of DAG 1998 scores by research groups. First, evaluations units were not recognised. Second, communication (frequency and amount) between research level and evaluation organisers (RMW and VSNU) was unsatisfactory. Third, the evaluation procedure took so much time that evaluation outcomes and recommendations were outdated as soon as they were published. In conclusion, it appears that the impact of external evaluations sometimes changes over time due to, for example, adjustments in the evaluation procedure.

Although attention to externally organised evaluations as well as knowledge accumulation in that area is extensive, none of the intermediary organisations ever conducted studies on the effects of external research evaluations. Consequently, intermediary organisations have little information on either positive or negative adjustments in internal work processes that are implemented as a response to the recommendations. In my opinion, this absence of follow-up studies is remarkable. It hampers communication between intermediaries and research institutes/groups. In this way it also fosters and increases the feeling of many researchers' that dealing with external evaluations is an administrative and bureaucratic burden. It can be argued that replacing the set of external research evaluations by one system of external research evaluation may facilitate and enlarge our knowledge and information about its long-term effects. In addition, it would improve communication between researchers and intermediaries, as will be further discussed in 7.5. The UK has some experience with one system of research evaluation (Research Assessment Exercise, RAE). Since 1986 research has been periodically assessed (Garret-Jones and Aylward, 2000); outcomes have direct implications for the amount of governmental research funding. Due to the RAE, centres of excellences have been developed and the performance of research has increased. Half of the research groups have improved their evaluation score compared to previous assessments (van de Meent and van Vliet, 2005). However, the improvement of performance also has a reverse side as studied by the British parliamentary committee. The rearrangement of research groups by assigning weak researchers to strong groups without negatively influencing the final assessment score and buying up excellent professors from competitive groups in order to count the publications of the transferred professor in the next RAE, are two examples of strategic games that are regularly played (van de Meent and van Vliet, 2005).

---

proposal.

		Intermediary level				Performance level			New national evaluation system
	Research indicators	ZonMw	RMW/ VSNU	ECOS	University	Non-university research institute	Medical faculty	Research group	KNAW/VSNU/ NWO and performance level[74]
Input	Financial resources	X	X	X	X	X	X	X	
	Hunan resources	X	X	X				X	X
	Research themes	X	X	X	X	X	X	X	
Throughput	Research management		X	X					X
	Collaboration		X	X					
	Training			X					
Output	Publications	X	X	X	X	X	X	X	X
	Citations				X	X	X	X	X
	Dissertations				X	X	X	X	
	Presentations							X	
	Awards	X							
	Innovation	X	X						X
	Patents					X	X		X
	Societal impact	X							X

Figure 13: Summary of research indicators used in Dutch medical and health research evaluations during the time empirical data for this thesis were being collected (2000-2002).

[74] Starting from 2003

Nowadays the RAE is a controversial but still functioning system within the UK. In conclusion, by taking the UK experiences into account, I would recommend the Dutch government and intermediaries frequently and adequately to assess the new national evaluation protocol and practices. These assessments will restrict the strategic games could be played. By doing so, the impact of these games would be kept to a minimum. Assessing the evaluation practices could result in (re)adjustment of the rules of national evaluation procedure. However, one should be aware that a system of rules will always in some way influence to some extent the behaviour of researchers. Strategic behaviour (also called goal shifting) is part of the game. However, I think that strategic games will have less impact in the Netherlands than in the UK, because the Dutch internal and external research evaluations are nowadays linked within one system.

7.3 Implications of internally organized research evaluations

The purpose of internally developed research evaluation practices is to assess staff members, research proposals and research output.

Research leaders generally evaluate staff members once a year. These job evaluations serve multiple purposes. They are not only used to assess staff but also to improve conditions in order to stimulate (individual) research performance. Specific evaluation practices to stimulate PhD students have been developed by a number of medical faculties and non-university research institutes.

Research proposals are generally pre-evaluated before they are submitted to external funding agencies. Pre-evaluations introduced by medical faculties and non-university institutes are organised by steering committees. Specific procedures for PhD research proposals have been developed. Which research proposals are pre-evaluated, varies between the faculties. It generally depends upon the proposal type (personal or programme grant) and the funding agency. Interestingly, non-university research institutes pre-evaluate any research proposal that is written. Their procedures are, compared to the ones that are used by medical faculties, more detailed. Furthermore, both research boards and research management teams are highly involved in pre-evaluations. Research groups also generally organise pre-evaluations of proposals. Evaluation outcomes are intended to assure performance. Research output evaluations are also prevalent. As shown in Figure 14, the number of publications and their SCI impact are the most frequently used indicators. However, the complexity of output evaluation practices varies between the medical faculties and non-university research institutes. Three types - none, restricted and extensive output evaluation practices- have been distinguished (see figure 11 in Chapter 5). The boundaries between them are not clear and sharp. From the interviews with research managers, it can be concluded that one evaluation type may gradually develop into another evaluation type over time.

Research output evaluations, which are internally organised, have a high impact on Dutch medical and health research. The majority of the responding medical and health research leaders have positive views about the outcome of internal research evaluations. They attach high value to evaluation outcome. In addition, research groups generally implement the recommendations that are given by peers. On faculty and institute level, boards use outcome to formulate and adjust their research strategy and profile. In addition, outcome is frequently used to re-adjust, develop, and (financially) stimulate local research programmes. These empirical results are in line with a study conducted by the American Sociological Association (2003). They found a few universities that use evaluation reports of departmental productivity to identify 'excellent' or potentially 'excellent' programmes and to justify the steering of additional resources to those departments.

7.4 Research management and performance of pre-clinical and clinical research groups

As shown in Figure 14 research throughput is only evaluated at the intermediary level. Improvement of research management, including leadership, to help research organisations, units and individual researchers to arrive at better decisions about future research, management and policy have recently become a new government goal. This renewed attention to research management results in an intensification of planning and managerial control within and outside academic research groups. We therefore analysed managerial control in chapter 6 of this thesis and studied its effects on the research performance of academic medical and health groups. The empirical study showed that there is only one internal research management activity which has a positive relationship with medical research performance in general. Offering special commendations to medical (both pre-clinical and clinical) research staff members, including non-financial prizes, in order to motivate them is positively correlated with the number of SCI publications, the percentage of externally obtained research funding and the NWO activity of research groups. In the regression analyses of contingencies and managerial control variables we found a positive linear relationship between special commendations and two medical research performance measures. These results are in line with suggestions by McKeachie (1979) and Omta (1997) who both argue that (immaterial) incentives reflect the competence achieved, and may stimulate staff and research groups to perform better.

The impact of the disciplinary setting is also studied in chapter 6 in order to assess whether such effects on performance are uniform. It was found that the disciplinary setting, the classification into pre-clinical and clinical research, has an impact on the relationships between managerial control and performance. As presented in Table 60, four differences have been found between pre-clinical and clinical research groups.

First, in clinical groups three internal management tools –organisation of research output evaluations; research policy meetings and research meetings- are important in predicting research performance. In pre-clinical groups the organisation of various internal research management activities shows relationships with particular research performance measures. No homogeneity was found among research management tools. In conclusion, managerial choices of leaders result in differences between performances measures. Different styles of research management are required to enable researchers to score well on the various research goals.

Second, the relationship between the time allocated to external research management activities and research performance is different for the two groups. In pre-clinical groups time allocated to external research management activities is important for obtaining research council involvement while in clinical groups time spent on these activities is important for predicting the number of SCI publications, and the DAG 1998 score. Third, views of pre-clinical leaders about the importance of collaboration with medical stakeholders are not linearly related to the research performance measures used in this study. In clinical groups, the collaboration with medical stakeholders generates relationships with different performance measures.

Finally, we have seen that contingencies are important predictors of research performance especially in pre-clinical groups. In clinical groups contingencies such as size and time allocation are the only important variables in getting papers published in SCI journals. Contingencies in these groups are not significantly correlated with performance measures concerning external research funding. Subsequently, no linear relationships were found.

In sum, it appears that the chosen subdivision of different types of research groups is a fruitful way of shedding light on the complex world of research management and research group performance. This study indicates that medical and health groups are different. Perhaps, this differentiation can, to some extent, be explained by the differences between pre-clinical and clinical research groups (as presented in Table 4 with regard to dependency on external financial resources. Unfortunately, we were not able to study the effects of managerial control on research performance in para-clinical research groups because a number of indicators were insufficient for measuring para-clinical research performance. For example, it turned out that most (72%) para-clinical research group leaders did not submit research proposals to the CMHR. Consequently, only a few groups received grants from the CMHR. Furthermore, only 23% of para-clinical research groups were assessed in the DAG 1998 evaluation. The number of para-clinical groups became too small to perform correlation and regression analyses. Additional in-depth and qualitative research is needed to analyse managerial control and its relationships with the research performance of para-clinical groups.

Table 60

Linear relationships between managerial control, contingencies and research performance of medical and health research groups: a summary

a. pre-clinical groups

	SCI Publications	**External funding**	**NWO activity**	**NWO rating**	**DAG 1998 score**
Internal control	+ Rewards: special commendations	+ Rewards: special commendations	+ Rewards: flexibility	+ Time to internal management	+ Research meetings
		+ Pre-evaluations of research proposals		- Research output evaluation	+&- Research output evaluations
				+ Research commitment	
				+ Research policy planning	
				- Direct communication	
External control			+ Time spent on external research activities		
Contingencies	+ Time for supervision	- Time for patient care	- Time for patient care		-Time for patient care
	- Work setting		+ Size		+ Time to research

b. clinical groups

	SCI Publications	**External funding**	**NWO activity**	**NWO rating**	**DAG 1998 score**
Internal control	+ Research meetings	+ Rewards: special commendations	+ Rewards: special commendations	+ Research meetings	
	+ Control of personnel resources	+ Research output evaluations	+ Research output evaluations	+ Research output evaluations	
	- Importance internal research evaluations		+ Research policy planning	+ Research policy planning	
External control	+ Time for external research activities	+ Time for external managerial control	- Importance of collaboration with clinic	-Importance of collaboration with clinic	+ Time for external management control
		+ Importance of collaboration with family doctors			+ Importance of collaboration with pharmaceutical industries
Contingencies	+ Size				
	+ Time for research				
	+ Time for patient care				

7.5 Discussion and some suggestions for further research

In this final section the possible contribution that the empirical results could make to management practice will be evaluated. Fruitful directions for further research and complementary follow up studies will be indicated and supported with literature.

An understanding of the determinants of medical and health research performance is a prerequisite for designing effective micro as well as macro research policy. It may give health research leaders and administrators tools to attract motivated individuals as well as to achieve organisational and project goals. Furthermore health research leaders and administrators may be stimulated to improve and control research group performance. This thesis revealed that different management actions may be advised for achieving a good overall group performance. Research leaders are recommended to compose their management team of staff members who have different tasks and also use different management styles. A challenge for future research is to gather data from junior and senior research staff members to get a more complete picture about the functioning of medical and health research groups (see the dotted line in figure 13). This future approach allows us to study the implications of managerial initiatives of research leaders such as for example the effects of leadership style. Over the years many authors have emphasised the style of leadership as an important determinant for obtaining performance. Studies have emphasised the importance of participative or consultative leadership in research groups (e.g. Pineau and Levy-Leboyer, 1983). Also positive relationships are found between transformational leadership and outcome variables of research project groups (e.g. Keller, 1995). However, studies concentrating on leadership and performance leave scope for other managerial control elements. We did some preliminary research (Heijnen, 2002) and found that transformational leadership has positive correlations with both the affective and the career-oriented commitment of medical staff members. Commitment has positive relationships with the functioning –measured by achievement orientation, competitiveness and persistence- of medical staff members. Also the culture and climate of research groups could be studied within future approaches. Although several scholars describe group culture and climate, studies focusing on the relationship between culture, climate and academic performance are rather limited. Within one decade, only two studies were published. Hurley (1995) studied the relationship between culture and innovative productivity in research groups. Once it has been established that the group has reliable structural properties, the innovativeness of the culture of this group may have a significant positive effect on innovative productivity. Personal and career development, and participative decision making explained 22% of the variation in the innovativeness of groups. In another study West and colleagues (1998) reported a relationship between the organisational climate and the excellence of departments, and vice versa. The strongest influences were indicated by the dimensions: ‘degree of formalisation’, ‘support for career development’ and ‘support for innovation’. We recently studied the effects of organisational culture and climate on the

performance of medical PhD students. A positive correlation between rule orientation and impact scores was found, suggesting that relatively low levels of autonomy may contribute to a researchers' productivity (Geerling et al., 2005). In future research a longitudinal study design is to be preferred. This might enable us to study the proposition that managerial activities and views actually enhance group performance and are not only initiated by leaders as a result of good performance. To assess the long-term effects of managerial activities within research groups, performance measures need to be measured at several points in time. Moreover, as we demonstrated in this study different performance measures show conflicting lessons for management.

The study of research staff members will also contribute to explaining how research group performance depends on the interaction between the leader and staff members, the so-called intra-group dynamics. Information about the internal social structure of research groups is essential for solving this research problem. Questions are for example: Are research management activities aimed at the group as a whole or at some staff members in particular? How does the group allocate tasks? What are the roles of staff members and research group leaders? Although research leaders and staff members have distinctive roles, it is essential -in improving research performance- that they develop some kind of mutual understanding in which both functions regard themselves as being representative of the research group. For example Shrum, Compalov and Genuth (2001) found that the organisation of interaction between structural components, such as research teams, is important. It is interesting to study these intra-group dynamics in the medical and health setting. Are there -in addition to differences in goals, activities and research management- also differences in intra-group dynamics between pre-clinical and clinical groups?

The issue of research management is still highly debated. This is also the case in the Dutch public media. Although most (90%) Dutch professors are generally satisfied with their jobs, they complain about low salaries, spending too much time on managerial tasks and unbearable bureaucracy (Broer, 2004). According to Borst (2004), these problems are generated by detailed government rules and the centralised way of managing universities by full-time managers who have limited knowledge about research and education practices. Borst proposes the *'meewerkend voorman'* model to solve this last problem. Research leaders should combine their part-time management job with running their own lab as is customary in the United States. Capable full-time research managers should assist research leaders, but should not be charged with making final decisions. In the case of research groups, this thesis showed that both internal and external managerial control has a positive impact on performance. Managing a research group actively makes a difference. However, an excellent researcher does not make an excellent group. I agree with Borst that an excellent creative leader supported by an enthusiastic research manager together could manage a research group. In my opinion, this research manager should not only accomplish administrative tasks (as suggested by Borst), but should also

coordinate research. He or she should be valued as an equal interlocutor and sparring partner for the research leader. That is why the research manager should have a university degree. For example a medical and health research group should seek a person who has a PhD in a relevant social science (e.g. science and technology studies or innovation processes) and a background in medicine or medical biology. Knowledge about and experience with evaluation systems is also necessary. As has been shown (e.g. van der Meulen, 1992; Omta, 1995; Spaapen, 1995, Wouters, 1999 and this thesis) research management and evaluation is a new but distinct sub-discipline within the social sciences. By using their particular skills research leaders and managers should define their own designated tasks and responsibilities. Generally, research leaders should focus on brainstorming and the elaboration of innovative research ideas, research communication, supervision and training of researchers. Research managers should have knowledge about and seek for opportunities concerning research funding, collaboration, training (including research schools) and research exchange. They also support the organisation of research evaluations and the implementation of evaluation outcomes and recommendations. Finally, they organise and coordinate internal research activities such as seminars and lectures. Research managers work at institutional level in a research institute or in an academic medical centre. Depending on the group size, research managers can support and coordinate a number of groups.

This thesis also showed that the introduction of the new national evaluation system –as developed by intermediaries and described in outline in chapter 4- for publicly funded research is a positive development for researchers. It will lower the administrative burden of procedures by linking external evaluations to internal ones. Furthermore, it will decrease researchers' work load because input and output data will not only be stored in an internal information system, which should be updated on an annual basis, but will also be presented to peers once-only in a standardised way. Another advantage is that it will improve communication among researchers because they will be able to ask colleagues for advice about the organisation and execution of self-evaluation. Finally, the introduction of the new research evaluation system will be an incentive for research leaders to improve further their internal evaluation system and the steering of research. As a consequence, the national comparability between medical and health research sub-disciplines, as presented in former national (DAG) evaluations, has disappeared. This is not harmful because, as described in this thesis, both research leaders and administrators reported many problems with the disciplinary evaluation procedure as well as with evaluation outcomes. Furthermore, the interactions between clinical and pre-clinical disciplines have strongly improved over the years. On the one hand, the increasing belief among medical and health researchers in the existence of a research continuum (*'from bench to bedside'*) has reduced the importance of sharp boundaries between the disciplines. On the other hand, this thesis shows that the medical and health sector is also not (yet) a coherent entity. It is conceivable that particular (clinical?) sub-disciplines attach value to evaluation and comparability on a national level. In this respect, the

recently accomplished bibliometric analysis of research conducted within Dutch university medical centres should be mentioned (Nederlandse Federatie van Universitair Medische Centra, 2004a). A comparison of the mean citation score of each UMC with the global citation score has indicated the ten most productive disciplines for each university medical centre. Oncology, cardiovascular diseases, biochemistry, molecular biology and haematology are disciplines that are prominently present in every UMC. Consequently, it is possible to compare these disciplines nationally.

The outcomes of this thesis are also of interest to health policymakers. The replacement of the external set of research evaluations by one system of external research assurance and the establishment of permanent internal evaluation of research reduces the participation and roles of intermediaries. However, in the first few years after introduction, intermediary organisations are essential for supporting and assisting researchers by putting the new evaluation system into practice. The establishment of Quality Assurance Netherlands Universities (QANU) in 2004 to support universities in the organisation of external assessments of research programmes and to advise on ways of improving internal quality assurance is a good development. However it should be kept in mind that university boards still have the primary responsibility for making choices with regard to evaluations of their institutes. Also the new role of the KNAW in developing knowledge about methodological issues concerning quality assurance (see KNAW, NWO and VSNU, 2001) and organising a meta-evaluation on the new evaluation process and its outcome (see VSNU, NWO and KNAW, 2003) is another good development. Finally, the accreditation and follow-up accreditation of research schools should be discussed. It could be hypothesized that the accreditation of research courses –instead of the accreditation of research schools– would be a logical future step within the framework of the accreditation of the Bachelors and Masters education systems. With regard to external input evaluations, the current approach of policymakers for enhancing research performance and stimulating new developments in research is to have groups develop detailed project proposals and to let these be judged by peers. In such procedures, limited attention is paid to managerial experience. Moreover, a strong emphasis on controlling and allocating scarce resources might lead to the neglect of the effects of organisational factors on the results achieved. Our results point to the need to develop a more specific set of assessment instruments for medical and health research. Also broadening incentives to include the public recognition of research work beyond direct funding might influence research direction and performance.

Finally, it would be interesting to study the relationship between managerial control and performance in other academic research fields. Can we identify managerial control variables, for example, special commendations and non-financial prizes, or others such as exposure in the media or discussions with users, which will stimulate the performance of academic research groups in general? And which managerial control variables are context-dependent? We believe that most parts of our questionnaire

are readily applicable to research activities in other fields. For example the concepts that together constitute the construct of managerial control refer to general processes such as control over personnel, resources and research processes and to communication and exchange with the environment of the research group. However, the questionnaire should also include specific questions concerning the chosen research field. To interpret the data well, detailed knowledge about the background, traditions and development within the chosen research field is required. A combination of qualitative (interviews) and quantitative (questionnaire) techniques is also recommended in future approaches.

KORTE NEDERLANDSE SAMENVATTING

Het proefschrift ‘Op zoek naar kwaliteit: onderzoeksmanagement in de Nederlandse publieke gezondheidssector’ geeft inzicht in de onderzoeksbeoordelingen waarmee academische onderzoeksgroepen werkzaam in de Nederlandse gezondheidssector te maken krijgen. Daarnaast zijn de effecten en implicaties van onderzoeksbeoordelingen onderzocht. Tenslotte is onderzoek gedaan naar de invloed van onderzoeksbeoordelingen en (andere) interne onderzoeksmanagement activiteiten op de performance van onderzoeksgroepen. Het onderzoek is uitgevoerd in de periode 1999-2004 en heeft een exploratief karakter. Data zijn verzameld met behulp van zowel kwalitatieve (interviews) als kwantitatieve (vragenlijst) technieken. In deze korte Nederlandse samenvatting zullen de opbrengsten van het proefschrift worden belicht.

Beoordeling van Nederlands gezondheidsonderzoek

Uit interviews met beleidsmedewerkers en onderzoeksleiders aangevuld met literatuur onderzoek wordt duidelijk dat onderzoeksgroepen te maken krijgen met verschillende soorten onderzoeksbeoordelingen in alle fases van het onderzoeksproces.

Externe beoordelingen van onderzoek worden in de regel uitgevoerd en/of gecoördineerd door intermediaire organisaties. Intermediaire organisaties zijn organisaties of onderdelen van organisaties die zowel gelieerd zijn aan de overheid als aan de wetenschappelijke gemeenschap. Voorbeelden zijn: (1) ZonMw, een fusieorganisatie van Zorgonderzoek Nederland (ZON) en Medische Wetenschappen NWO (NWO-MW). Deze organisatie is verantwoordelijk voor het stimuleren van gezondheidsonderzoek, dat zich uitstrekt van fundamenteel onderzoek tot uitvoeringsprojecten in de praktijk. (2) de Raad voor de Medische Wetenschappen van de KNAW (RMW-KNAW) tot voor kort verantwoordelijk voor het uitbrengen van het Discipline Advies Geneeskunde. Dit was een vijfjaarlijkse beoordeling van al het gezondheidsonderzoek dat binnen medische faculteiten/clusters en para-universitaire instituten had plaatsgevonden. (3) De Vereniging van Samenwerkende Nederlandse Universiteiten (VSNU), organisator van beoordelingen van de universitaire onderzoeksprogramma’s met behulp van internationale visitatie commissies. In het kader van het gezondheidsonderzoek heeft men vooral in de beoordeling van 1998 nauw samengewerkt met de KNAW. (4) Raad voor Gezondheidsonderzoek, geeft advies aan ministeries over prioriteiten in het gezondheidsonderzoek en de technologieontwikkeling in deze sector, evenals de daarbij behorende infrastructuur. (5) Erkenningscommissie Onderzoekscholen, een zelfstandige commissie verbonden aan de KNAW en verantwoordelijk voor de erkenning en hererkenning van onderzoekscholen.

Onderzoeksleiders werkzaam in de gezondheidssector geven aan dat het onderzoek ook binnen de eigen organisatie wordt beoordeeld. Deze beoordelingen worden georganiseerd door universiteiten, medische faculteiten en onderzoeksinstituten. Daarnaast organiseren veel onderzoeksgroepen zelf ook beoordelingen om zo de gewenste onderzoeksprestaties te bereiken. Zo worden onderzoeksvoorstellen

intern op verschillende organisatieniveaus beoordeeld voordat indiening bij externe financiers plaatsvindt. Ook zijn er speciale procedures voor PhD onderzoeksvoorstellen ontwikkeld. Welke onderzoeksvoorstellen worden beoordeeld verschilt en hangt af van het type onderzoeksvoorstel en onderzoeksfinancier. Ook de onderzoeksoutput wordt op verschillende organisatie niveaus intern beoordeeld. De complexiteit van de beoordelingsprocedures verschilt sterk. Het aantal publicaties en de SCI impact zijn de meest gehanteerde indicatoren in de interne output beoordelingen.

De bovengenoemde externe en interne onderzoeksbeoordelingen zijn allen (in meer of mindere mate) gebaseerd op peer review. Peers geven commentaar op voorgesteld en uitgevoerd onderzoek, daarnaast geven zij aanbevelingen om onderzoek te verbeteren. Bij het beoordelen van onderzoeksvoorstellen geven peers daarnaast ook nog advies over de mogelijke financieringskans. De inhoud van het protocol verschilt per onderzoeksbeoordeling. De selectie van peers, de afweging van de onafhankelijkheid en deskundigheid van peers, de toepassing van wederhoor, het gebruik van de rankingsmethode en de afweging om kwantitatieve gegevens te verzamelen en te gebruiken in het beoordelingsproces zijn enkele onderwerpen die in het protocol vermeld worden.

In interne onderzoeksbeoordeling worden directe collega's en/of onderzoekers werkzaam in dezelfde instelling gevraagd als peers. Externe beoordelingen maken gebruik van peers die volkomen onafhankelijk zijn van het te beoordelen evaluatie object. Bij het beoordelen van wetenschappelijke kwaliteit van onderzoek zijn peers altijd wetenschappers.

De maatschappelijke kwaliteit en impact van gezondheidsonderzoek wordt in Nederland steeds belangrijker gevonden, vooral binnen externe onderzoeksbeoordelingen. Methodes en indicatoren om maatschappelijke aspecten van gezondheidsonderzoek te meten worden momenteel ontwikkeld. Dit zijn voornamelijk kwantitatieve tools die gebruikt kunnen worden binnen zelfevaluaties en externe audits. Daarnaast zijn er (KNAW) voorstellen gedaan om het traditionele peer review proces uit te breiden met stakeholders zoals patiënten organisaties.

In de praktijk blijkt dat de afstemming in de vorm van planning, uitvoering en informatievraag tussen de externe onderzoeksbeoordelingen minimaal is. Dit is opmerkelijk en draagt bij aan een hoge beoordelingslast per onderzoekseenheid. Dit komt onder meer tot uitdrukking in veelvuldige verzoeken aan onderzoekers om informatie voor externe beoordelingen te leveren, vaak juist weer net in een andere vorm. Om deze problemen te minimaliseren heeft de werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek in 1999 een nieuw beoordelingsysteem voor academisch onderzoek ontwikkeld. Dit is verder uitgewerkt en heeft in 2003 geleid tot de publicatie van het standaard evaluatie protocol. Het evaluatieprotocol verbindt de externe met de interne kwaliteitszorg met het doel de bureaucratisering te verminderen. Onderzoeksgroepen initiëren zelfevaluaties om zodoende de aansturing van onderzoeksinstituten te verbeteren. Tevens zijn de zelfevaluaties het startpunt voor zesjaarlijkse externe evaluaties. De externe evaluaties zijn bedoeld onderzoeksorganisaties, managers en individuele onderzoekers te helpen om (betere) beslissingen te nemen over de onderzoeksplannen,

onderzoeksmanagement en –beleid. Anno 2006 wordt dit evaluatie protocol in elk universitair medisch centrum gebruikt.

Implicaties van beoordelingen van gezondheidsonderzoek

Nu een overzicht is gegeven van de beoordelingen waar onderzoeksgroepen mee te maken krijgen is het interessant te onderzoeken welke impact de beoordelingen hebben op het gezondheidsonderzoek. Hiertoe zijn open interviews gehouden met onderzoeksleiders, onderzoeksmanagers werkzaam op facultair niveau en beleidsmedewerkers werkzaam bij intermediaire organisaties. Deze interviews laten zien dat de impact van extern georganiseerde onderzoeksbeoordelingen sterk verschillen. De beoordeling van onderzoeksvoorstellen door ZonMw heeft veel impact op het gezondheidsonderzoek. Aspecten van de ZonMw beoordelingsprocedures zijn overgenomen en gebruikt zijn in interne ex-ante onderzoeksbeoordelingen. Ook het verkrijgen van een ZonMw subsidie staat hoog aangeschreven in het onderzoeksveld. Bovendien worden de ZonMw beoordelingsuitkomsten in sommige medische faculteiten gebruikt als indicator bij de verdeling van interne onderzoeksbudgetten. De impact van de accreditatie van onderzoekscholen is de laatste jaren sterk afgenomen, vooral door de sterke toename in het aantal dat accreditatie verkregen heeft. Ook het Discipline Advies Geneeskunde heeft weinig impact op het onderzoeksveld gehad. Onderzoeksgroepen ondervonden problemen met de evaluatie procedure, zoals de geringe communicatie met de organisatoren. Ook nam de evaluatie procedure zoveel tijd in beslag dat de uitkomsten verouderd waren op het moment dat de resultaten openbaar werden gemaakt. De intermediaire organisaties die in deze studies onderzocht zijn hebben zelf geen (follow-up) onderzoek gedaan naar de effecten en impact van hun onderzoeksbeoordeling(en) op het veld. Dit is opmerkelijk gezien de veelvuldige aandacht voor beoordelingen en de grote hoeveelheid kennis die is en nog wordt verzameld. Het ontbreken van follow-up studies doet de communicatie tussen de intermediaire en onderzoekers geen goed. Het versterkt het gevoel te maken te hebben met een bureaucratische en administratieve last. Het nieuwe evaluatie protocol dat de interne met de externe kwaliteitszorg verbindt zal in de toekomst informatie en kennis genereren over de lange termijn effecten van onderzoeksbeoordelingen. Hierdoor zal ook de communicatie tussen intermediairen en onderzoekers versterkt worden.

Interviews en survey uitkomsten laten zien dat interne output evaluaties veel impact hebben op het Nederlandse gezondheidsonderzoek. In de meeste medische faculteiten en onderzoeksinstituten hebben interne evaluatie uitkomsten en aanbevelingen directe implicaties voor het lokale onderzoeksbeleid. Interne evaluatie resultaten worden gebruikt bij het formuleren en aanpassen van het onderzoeksprofiel en de onderzoekstrategie. Bovendien worden de resultaten gebruikt bij het ontwikkelen, sturen en (financieel) stimuleren van locale onderzoeksprogramma's. In twee medische faculteiten wordt op basis van de uitkomsten van interne onderzoeksbeoordelingen een gedeelte van het onderzoeksgeld verdeeld. De meeste onderzoeksleiders zijn positief over de uitkomsten van interne onderzoeksbeoordelingen. Ze hechten veel waarde aan de evaluatie uitkomsten. Over het

algemeen implementeren de meeste onderzoeksleiders de aanbevelingen die peers aanreiken. De mate waarin evaluatie uitkomsten en aanbevelingen geïmplementeerd zijn in de onderzoeksgroep wordt regelmatig gebruikt als indicator in een volgende onderzoeksoutput evaluatie.

Onderzoeksmanagement van gezondheidsonderzoek

Een voorwaarde voor het ontwerpen van effectief onderzoeksbeleid, op zowel micro als macro niveau, is het hebben van inzicht in de determinanten van performance. Dit geeft onderzoeksleiders en beleidsmedewerkers tools om gemotiveerde onderzoekers aan te trekken en om organisatie en project doelen te behalen. Bovendien worden zij zo gestimuleerd om de performance van de onderzoeksgroep te controleren en te verbeteren. Ook is de relatie tussen het hanteren van managementactiviteiten/visies en performance van groepen onderzocht. Een vragenlijst is door 160 onderzoeksleiders werkzaam in het gezondheidsonderzoek beantwoord. De kwantitatieve studie laat zien dat een onderzoeksklimaat waarin waardering wordt uitgesproken en zichtbaar wordt gemaakt door middel van het verstrekken van eervolle vermeldingen en niet-financiële prijzen als stimulans voor onderzoekers een positieve relatie heeft met drie performance maten (aantal SCI publicaties, percentage extern verkregen onderzoeksfinanciering en mate van NWO activiteit). De invloed van disciplinaire achtergrond is ook onderzocht. De gereageerde onderzoeksgroepen zijn onderverdeeld in twee categorieën, te weten pre-klinische en klinische groepen. Uit onderzoek blijkt dat de disciplinaire achtergrond impact heeft op de relatie tussen (intern en extern) onderzoeksmanagement en de performance van groepen. Er zijn vier verschillen gevonden tussen pre-klinische en klinische groepen. Tabel 61 geeft een samenvattend overzicht.

1. In klinische groepen hebben drie management activiteiten, te weten het organiseren van output evaluaties, onderzoeksbijeenkomsten en discussies over onderzoeksbeleid, een positieve relatie met performance. Echter in pre-klinische groepen leiden de management keuzes die onderzoeksleiders maken tot verschillen tussen performance maten. Het verdient dan ook de aanbeveling verschillende management stijlen te hanteren om zodoende een onderzoeksgroep op verscheidene performance maten goed te doen scoren. Onderzoeksleiders worden dan ook geadviseerd om hun stafleden verschillende taken te geven en elk een andere management stijl te hanteren.
2. De relatie tussen de tijd die de onderzoeksleider besteed aan externe management activiteiten en performance is verschillend voor de twee soorten onderzoeksgroepen. In pre-klinische groepen is extern onderzoeksmanagement belangrijk bij het indienen van ZonMw en NWO onderzoeksvoorstellen en het verkrijgen van grants. In klinische groepen is tijdsbesteding aan extern onderzoeksmanagement belangrijk bij het voorspellen van SCI publicaties en de hoogte van de DAG 1998 score.
3. De meningen van pre-klinische onderzoeksleiders met betrekking tot het belang van samenwerking met medische stakeholders zijn niet lineair gerelateerd aan de onderzoeksperformance maten die

gebruikt zijn in deze studie. In klinische groepen bestaat er wel een (zowel positieve als negatieve) lineaire relatie met verschillende performance maten.

4. Tenslotte heeft de disciplinaire achtergrond ook invloed op de relatie tussen contingentie factoren en de performance van groepen. Contingentie variabelen zijn vooral in pre-klinische onderzoeksgroepen belangrijke predicatoren voor performance.

Vervolg onderzoek

Om een compleet plaatje te krijgen van het functioneren van onderzoeksgroepen (werkzaam in de gezondheidsector) is het belangrijk data te verzamelen van zowel junior als senior stafleden. Deze aanpak stelt ons ook in staat de gevolgen de management initiatieven van onderzoeksleiders op de performance van groepen te onderzoeken. Onderzoek naar het effect van leiderschapstijlen en onderzoek naar de cultuur en het klimaat van onderzoeksgroepen zijn enkele voorbeelden.

Het betrekken van stafleden in het onderzoek zal ook inzicht geven in de interactie tussen de onderzoeksleider en stafleden. Interessant is om te onderzoeken wat de invloed van deze interactie op de groepsperformance is.

In vervolg onderzoek zal een longitudinale aanpak gekozen worden. Dit stelt ons in staat de hypothese dat management activiteiten en visies met betrekking tot onderzoeksmanagement daadwerkelijk de performance van onderzoeksgroepen verhogen en niet geïnitieerd worden door onderzoeksleiders als gevolg van goede performance. Om de lange termijn effecten van management activiteiten te bestuderen dient groepsperformance op meerdere momenten in tijd te worden gemeten.

Tenslotte is het interessant de relatie tussen het hanteren van managementactiviteiten/visies en performance ook in andere academische sectoren te onderzoeken. Bestaan er management variabelen die de performance van academische onderzoeksgroepen in het algemeen stimuleren? Het verstrekken van eervolle vermeldingen en niet-financiële prijzen door onderzoeksleiders, het bediscussiëren van onderzoek met gebruikers en de mate van publiciteit in de media of zijn enkele voorbeelden die nader onderzocht kunnen worden.

Tabel 61

Lineaire relaties tussen onderzoeksmanagement, contingenties en performance maten van academische groepen werkzaam in de gezondheidsector

a. pre-klinische groepen

	SCI Publicaties	**Externe financiering**	**NWO activiteit**	**Status bij NWO**	**DAG 1998 score**
Interne onderzoeksmangement variabelen	+ eervolle vermeldingen en niet-financiële prijzen	+ eervolle vermeldingen en niet-financiële prijzen + beoordelen van onderzoeksvoorstellen	+ flexibele werkuren	+ tijd aan intern onderzoeksmanagement - beoordelen van onderzoeksoutput + betrokkenheid bij onderzoek + organiseren van bijeenkomsten over onderzoeksbeleid - Directe communicatie	+ organiseren van onderzoeksbijeenkomsten +&- beoordelen van onderzoeksoutput
Externe onderzoeksmanagement variabelen			+ tijd aan externe onderzoeksactiviteiten		
Contingentie variabelen	+ tijd aan begeleiding van onderzoekers - werkomgeving	- tijd aan patiëntenzorg	- tijd aan patiëntenzorg + Groepsgrootte		- tijd aan patiëntenzorg + tijd aan onderzoek

b. klinische groepen

	SCI Publicaties	**Externe financiering**	**NWO activiteit**	**Status bij NWO**	**DAG 1998 score**
Interne onderzoeksmanagement variabelen	+ organiseren van onderzoeksbijeenkomsten + voldoende personeel - belangrijkheid van interne onderzoeksevaluaties	+ eervolle vermeldingen en niet-financiële prijzen + beoordelen van onderzoeksoutput	+ eervolle vermeldingen en niet-financiële prijzen + beoordelen van onderzoeksoutput + organiseren van bijeenkomsten over onderzoeksbeleid	+ organiseren van onderzoeksbijeenkomsten + beoordelen van onderzoeksoutput + organiseren van bijeenkomsten over onderzoeksbeleid	
Externe onderzoeksmanagement variabelen	+ tijd aan externe onderzoeksactiviteiten	+ tijd aan externe management activiteiten + belangrijkheid van samenwerking met huisartsen	- belangrijkheid van samenwerking met de kliniek	- belangrijkheid van samenwerking met de kliniek	+ tijd aan externe management activiteiten + belangrijkheid van samenwerking met farmaceutische industrie
Contingentie variabelen	+ groepsgrootte + tijd aan onderzoek + tijd aan patiëntenzorg				

LITERATURE

Allen, R. S., & Helms, M. M. (2001). Reward practices and organizational performance. *Compensation and Benefits Review, 33*(4), 74-85.

Allen, T. J. (1970). Communication networks in R&D laboratories. *R and D Management, 1*(1), 14-21.

Allison, P. D., & Long, J. S. (1990). Departmental effects on scientific productivity. *American Sociological Review, 55* (4), 469-478.

American Sociological Association. (2003). *Task force on the implications of the evaluation of faculty productvity and teaching effectiveness*. Washington DC: American Sociological Association.

Andrews, F. (1979a). Motivation, diversity and performance. In F. Andrews (Ed.), *Scientific productivity: the effectiveness of research groups in six countries* (pp 253-289). Cambridge: Cambridge University Press/UNESCO.

Andrews, F. (1979b). *Scientific productivity: the effectiveness of research groups in six countries*. Cambridge: Cambridge University Press/UNESCO.

Badawy, M. K. (1988). Managing human resources. *Research Technology Management, 31*(5), 19-35.

Baird, L. (1986). What characterizes a productive research department? *Research Higher Education, 25*(3), 211-225.

Bartelse, J. (1999). *Concentrating the minds: the institutionalisation of the graduate school innovation in Dutch and German higher education.* University of Twente, Enschede.

Bazeley, P. (1998). Peer review and panel decisions in the assessment of Australian research council project grant applicants: what counts in a highly competitive context? *Higher Education, 35* (4), 435-452.

Begeleidingscommissie Experimentele Visitaties Gezondheidsonderzoek. (1994). *Kwaliteit verzekerd: Advies Visitaties Gezondheidsonderzoek.* Amsterdam: KNAW.

Bingham, C. (1999). Peer review on the internet: are there faster, fairer, more effective methods of peer review. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (pp. 205-224). London: BMJ Books.

Bland, C. J., & Ruffin, M. T. (1992). Characteristics of a productive research environment: Literature Review. *Academic Medicine, 67*(6), 385-397.

Blume, S. S., & Geesink, I. (2000). Vaccinology: an industrial science? *Science as Culture, 9*(1). 41-72

Blume, S. S., & Sinclair, R. (1973a). Chemists in British univerisities: a study of the reward system in science. *American Sociological Review, 38*(1), 126-138.

Blume, S. S., & Sinclair, R. (1973b). *Research environment and performance in British Univerisity Chemistry* (Science Policy Studies no. 6). London: Her majesty's stationery office.

Boor, M. (1982). The citation impact factor-another dubious index of journal quality. *American Psychologist, 37*(2), 975-977.

Borst, P. (2004). Onder professoren. *NRC.*

Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis II: dynamical aspects. *Journal of the American Society for Information Science, 42*(4), 252-266.

Broer, T. (2004). Nederland heeft een academisch minderwaardigheidscomplex; Buitenlandse professoren over de Nederlandse universiteiten. *Vrij Nederland*.

Buruma, O. J. S. (1997). De positie van de academische medische clusters in de 21-ste eeuw. *Medisch Contact, 52*(27-28), 864-866.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. *Social Science Information, 22*(2), 191-235.

Callon, M., Law, J., & Rip, A. (1986). *Mapping the dynamics of science and technology*. London: MacMillian Press Ltd.

Chawla, A., & Singh, J. P. (1998). Organizational environment and performance of research groups - a typological analysis. *Scientometrics, 43*(3), 373-391.

Cheung, S. F. (2003). *Doctoraal scriptie: Kwaliteitsbeoordeling wetenschappelijk gezondheidsonderzoek uitgevoerd door fondsenwervende instellingen*. Amsterdam: Vrije Universiteit.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless science. Peer review and U.S. Science Policy*. Albany, New York: SUNY Press.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submission: a cross-disciplinary investigation. *Behavorial and Brain Sciences,14*(1), 119-186.

Clark, B. R. (1983). *The higher education system. Academic organization in cross-national perspective*. Berkeley: University of California Press.

Clark, B. R. (1995). *Places of Inquiry: Research and advanced education in modern universities*. Berkeley and Los Angeles: University of California Press.

Cohen, B. P., Kruse, R. J., & Anbar, M. (1982). The social structure of scientific research teams. *Pacific Sociological Review, 25*(2), 205-232.

Cole, J. R., & Cole, S. (1971). Measuring the quality of sociological research: problems in the use of the Science Citation Index. *American Sociologist, 6(issue February)*, 23-29.

Cole, J. R., & Cole, S. (1973). *Social stratification of science*. Chicago: The University of Chicago Press.

Cole, J. R., & Zuckerman, H. (1984). The productivity puzzle: persistence and change in patterns of publication of men and women scientists. In M.W. Steinkempt & M.L. Maehr (Eds.), Advances in motivation and achievement (pp. 217-258),vol. 2. Greenwich, Connecticut: JAI Press INC

Cole, S. (1998). How does peer review works and how can it be improved? *Minerva, 36*(4), 179-189.

Cole, S., & Cole, J. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review, 32*(3), 377-390.

Cole, S., Rubin, L., & Cole, J. R. (1978). *Peer review in the National Science Foundation: phase one of a study*. Washington: National Academy of Sciences.

Converse, J., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaires*. Thousand Oaks, CA: Sage publications, Inc.

Council for Medical Sciences. (2002). *The societal impact of applied health research: towards a quality assessment system*. Amsterdam: KNAW.

Cozzens, S. E., Healey, P., Rip, A., & Ziman, J. (1990). *The research system in transition*. Dordrecht: Kluwer Academic Publishers.

Crane, D. (1965). Scientists at major and minor universities: a study of productivity and recognition. *American Sociological Review, 30*(5), 699-714.

Creswell, J. W. (1985). *Faculty research performance: lessons from the sciences and social sciences*. Washington D.C.: Association for the study of higher education.

Culpepper, L., & Franks, P. F. (1983). Family medicine research: status at the end of the first decade. *JAMA, 249*(1), 63-68.

Daniel, H. D. (1998). *Guardians of science: fairness and reliability of peer review*: Weinheim.

Dijstelbloem, H., Spaapen, J., & Wamelink, F. (2002). *Knocking at the gatekeepers door: the distribution of accountability in pharmaceutical research.* Paper presented at the Society for Social Studies of Science (4S) conference, 26-th annual meeting, Milwaukee, USA.

Dill, D. D. (1985). Theory versus practice in the staffing of R&D laboratories. *R&D Management, 15*(3), 227-241.

Dillman, D. (2000). *Mail and internet surveys: the tailored design method*. New York: Wiley.

Drongelen, K.-v., & Bilderbeek, J. (1999). R&D performance measurement: more than choosing a set of metrics. *R&D Management, 29*(1), 35-46.

Edwards, R. G., Van Steirteghem, A. C., & Richardson, M. J. (1993). Impact factors-more misleading than ever. *Human Reproduction, 8*(7), 983-985.

Efferth, T. (2001). Quality management for biomedical research in Germany. *Theory in Biosciences, 120*(1), 57-63.

Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics. Quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.

Ellenbroek, S. P. H., Van Ark, G., & Klasen, E. C. (2002). Vergelijking van de uitgaven aan gezondheidsonderzoek in 7 westerse landen in 1997: Nederland op de laatste plaats. *Nederlands Tijdschrift voor Geneeskunde, 146*(29), 1369-1374.

Endler, N. S. (1978). Beyond citation counts: developing reseach profiles. *Canadian Psychological Review, 19*(2), 152-157.

Etzkowitz, H. (2003). Resarch groups as 'quasi-firms': the invention of the entrepreneurial university. *Research Policy, 32*(1), 109-121.

Etzkowitz, H., & Leydesdorff, L. (1995). The triple helix of university-industry-government relations: a laboratory for knowledge based economic developments. *EASST Review, 14*(1), 11-19.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and 'Mode2' to a triple helix of university-industry-government relations. *Research Policy, 29*(2), 109-123.

European Commission. (1997). *Second European Report on S&T Indicators*. Luxembourg: Office for Official Publications of the European Communities.

Fairweather, J. S. (1993). Academic values and faculty rewards. *Review of Higher Education, 17*(1), 43-68.

Feist, G. J., & Gorman, M. E. (1998). The psychology of science: Review and integration of a nascent discipline. *Review of GeneralPsychology, 2*(1), 3-47.

Fenger, P. (1992). Research councils; buffers under cross-pressures. *Research notes from the Netherlands, 1*, 2-7.

Fenger, P., & Frankfort, J. G. (1989). New structures in postgraduate research training in the Netherlands. *Higher Education, 1*(1), 20-35.

Fletcher, R. H., & Fletcher, S. W. (1999). The effectiveness of editorial peer review. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (pp. 45-57). London: BMJ Books.

Fowler, J. (1993). *Survey research methods* (Vol. 1). Newbury Park: Sage Publications.

Fuhrer, M. J., & Grabois, M. (1985). Grant application and review procedures of the National Institute of handicapped research: survey of applicant and peer review opinion. *Archives of Physical Medical Rehabilitation, 66*(5), 318-321.

Garret-Jones, S., & Aylward, D. (2000). Some recent developments in the evaluation of university research outcomes in the United Kingdom. *Research Evaluation, 8*(1), 69-75.

Geerling, M., Gilder, D., Groenewegen, P., & Van der Weijden, I. (2005). Organizational Culture its Effects on Performance and Career Choices in Medical Research. *Research in Higher Education, under review*.

Gibbons, M., Limoges, C., Nowotny, G. H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. London, Thousand Oaks, New Dehli: SAGE Publicaitons.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibiometric research. *Scientometrics, 53*(2), 171-193.

Goldbeck-Wood, S. (1999). Evidence on peer review-scientific quality control or smokescreen? *British Medical Journal, 318*(7175), 44-45.

Granovetter, M. (1973). The strength of weak ties. *American Sociological Review, 78*(6), 1360-1380.

Groenewegen, P. (1988). *Scientists, audiences and resources.* University of Amsterdam, Amsterdam.

Gustad, J. W. (1960). *The career decisions of college teachers*. Washington DC.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (fifth ed.). New Jersey: Prentice Hall.

Hansson, S. (1995). Impact factor as a misleading tool in evaluation of medical journals. *Lancet, 346*(8979), 906.

Harnad, S. (1996). Implementing peer review on the net: scientific quality control in scholarly electronic publications. In R. P. Peek & G. B. Newby (Eds.), *Scholarly publishing. The electronic frontier* (pp. 103-118). Cambridge, MA: MIT Press.

Harvey, J., Pettigrew, A., & Ferlie, E. (2002). The determinants of research group performance: towards mode 2? *Journal of Management Studies, 39*(6), 747-774.

Hazeu, C. A. (1990). *Science policy: A behavioural approach*: Alderschot: Averbury/Gowering Publishing.

Heijnen, A. (2002). *Doctoraalscriptie: Stijl van leidinggeven, betrokkenheid, het functioneren van onderzoeksleden en de kwaliteit van gezondheidsonderzoek.* Amsterdam: Vrije Universiteit.

Hicks, D. (1987). Limitations of co-citation analysis as a tool for science policy. *Social Studies of Science, 17*(2), 295-316.

Hinze, S. (1997). *Mapping of structures in science and technology: bibliometric analyses for policy purposes.* Karlsruhe: Fraunhofer ISI

Horrobin, D. F. (1990). The philosophical basis of peer review and the supression of innovation. *Journal of the American Medical Association, 263*(10), 1438-1441.

Horton, B. (1999). From bench to bedside...research makes the translational transition. *Nature, 402*(6758), 213-215.

Hurley, R. F. (1995). Group culture and its effects on innovative productivity. *Journal of Engineering and Technology Management, 12(1-2)*, 57-75.

Husso, K., Karjaleinen, S., & Parkkari, T. (2000). *The state and quality of scientific research in Finland: A review of scientific research and its environment in the late 1990s*: Akademy of Finland.

Irvine, J., & Martin, B. R. (1984). *Foresight in science: picking the winners*. London and Dover: Frances Pinter.

Jain, R. K., & Traindis, H. C. (1990). *Management of research and development organizations: managing the unmangeable*. New York: John Wiley & Sons.

Jenskins, G. D., Gupta, N., Mitra, A., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of emperical research. *Journal of Applied Psychology, 83*(5), 777-787.

Jordan, J. M., Meador, M., & Walter, S. (1988). Effects of department size and organization on the research productivity of academic economics. *Economics of Education Review, 7*(2), 251-255.

Keller, R. T. (1995). Transformational leaders make a difference. *Research Technology Management, 18*(3), 41-44.

King, J. (1987). A review of bibliometric and other science indicators in their role in research evaluation. *Journal of Information Science, 13*, 261-276.

Klasen, E. C. (2000). Management van gezondheidsonderzoek. Leiden: Leiden Universiteit.

Klasen, E. C., & Overbeke, A. J. P. M. (2002). Melden van wetenschappelijk wangedrag in gezondheidsonderzoek. *Nederlands Tijdschrift voor Geneeskunde, 146*(35), 1622-1624.

Klazinga, N. S., & Van Everdingen, J. J. E. (2001). Het referentenrapport. *Mediator, 12*(3), 35-36.

KNAW (1988). *Disciplinepan Geneeskunde: samenhang, samenwerking en profilering*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen, Commissie Geneeskunde, Subcommissie integratie disciplineplan geneeskunde.

KNAW (1994). *Discipline-advies geneeskunde 1994*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen, commissie geneeskunde, subcommissie ten behoeve van het discipline-advies geneeskunde.

KNAW (1995). *Erkenning Onderzoekscholen*. Amsterdam: KNAW.

KNAW (1999). *Discipline report on biomedical and health sciences research in the Netherlands 1998*. Amsterdam:KNAW.

KNAW (2001). *Nota van toelichting op wijziging in 2001 van protocol Erkenningsprocedure Onderzoekscholen*. Amsterdam: KNAW.

KNAW (2002). *Erkenning Onderzoekscholen*. Amsterdam: KNAW.

KNAW, NWO, & VSNU. (2001). *Kwaliteit Verplicht: naar een stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*. Amsterdam, Den Haag, Utrecht: KNAW, NWO, VSNU.

Knorr, K. D., Mittermeir, R., Aichholzer, G., & Waller, G. (1979). Individual publication productivity as a social position effect in academic and industrial research units. In F. M. Andrews (Ed.), *Scientific productivity: the effectiveness of research groups in six countries* (pp 55-94). Cambridge: Cambridge University Press.

Koenen, L. (2000). Wetenschappelijk onderzoek, Dilemma's en verleidingen. *Akademie Nieuws*, 12-15.

Kyvik, S., & Marheim Larsen, I. (1994). International contact and research performance. *Scientometrics, 29*(1), 161-172.

Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*. Reading: Addison-Wesley.

Lewis, D. R., & Becker, W. E. (1979). *Academic rewards in higher education*. Cambridge Massachussetts: Ballinger.

Lewison, G. (2002). Researchers' and users' perception of the relative standing of biomedical papers in different journals. *Scientometrics, 53*(2), 229-240.

Long, J. S. (1978). Productivity and academic position in the scientific career. *American Sociological Review, 43*(6), 899-908.

Long, J. S., & McGinnis, R. (1981). Organizational context and scientific productivity. *American Sociological Review, 46*(4), 422-442.

Louis, K. S., Blumenthal, D., Gluck, M. E., & Stoto, M. A. (1989). Entrepreneurs in Academe: an exploration of behaviors among life scientists. *Administrative Science Quarterly, 34*(1), 110-131.

Lyall, C., Bruce, A., Firn, J., Firn, M., & Tait, J. (2004). Assessing end-use relevance of public sector research organisations. *Research Policy, 33*(1), 73-87.

Martin, J. (1998). *Organizational behaviour*. London: ITIP.

McKeachie, W. (1979). Perspectives from Psychology. In D. Lewis & W. Becker (Eds.), *Academic rewards in higher education*. Massachusetts: Ballinger.

Medical Committee of the Royal Netherlands Academy of Arts and Sciences. (2002). *The societal impact of applied health research: towards a quality assessment system* (draft report No. CG01/GZO/123). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Meijerink, R. (1999). Een gulden eeuw voor universitair onderwijs? *Tijdschrift voor Bestuurskunde, 8*(8), 377-384.

Meltzer, L. (1956). Scientific productivity in organizational settings. *Journal of social issues, 12*, 32-40.

Merton, K. M. (1973). *The sociology of science: theoretical and empirical investigations*. Chicago/London: The university of Chicago press.

Meyer, M. (2003). Academic entrepreneurs or entrepreneurial academics? Research-based ventures and public support mechanisms. *R&D Management, 33*(2), 107-115.

Ministerie van Onderwijs Cultuur en Wetenschappen. (1996). *Topkwaliteit en relevantie: samenvatting van het wetenschapsbudget 1997*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.

Ministerie van Onderwijs Cultuur en Wetenschappen. (1999). *Wie oogsten wil, moet zaaien: wetenschapsbudget 2000*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1983). *On the measurement of research performance*. Leiden: Science Studies Unit.

Moed, H. F., & Van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature, 381*(6579), 186.

Moorhead, B., & Griffin, R. W. (1998). *Organizational behaviour: managing people in organizations*. New York: Houghton Mifflin Company.

Morris, N. (2000). Science policy in action: policy and the researcher. *Minerva, 38*(4), 425-451.

Morris, N. (2002). The developing role of departments. *Research Policy, 31*(5), 817-833.

Moxham, H., & Anderson, J. (1992). Peer review. A view from the inside. *Science and Technology Policy, 5*(1), 7-15.

National Science Board. (1998). *Science and Engineering Indicators 1998.* Arlington: National Science Foundation.

Nederhof, A. J., Van Leeuwen, T. N., & Visser, M. S. (1998a). *A bibliometric analysis of the faculty of medicine, Univerisity of Groningen: A decade of research*. Leiden: Center for Science and Technology Studies.

Nederhof, A. J., Van Leeuwen, T. N., & Visser, M. S. (1998b). *A bibliometric analysis of the Netherlands school of primary care research.* Leiden: Center for Science and Technology Studies.

Nederhof, A. J., & Van Raan, A. F. J. (1993). A bibliometric analysis of six economics research groups: a comparison with peer review. *Research Policy, 22*(4), 353-368.

Nederhof, A. J., & Visser, M. S. (1997). *A bibliometric study of the cardiovascular research institute Maastricht (CARIM)*. Leiden: Center for Science and Technology Studies.

Nederlandse Federatie van Universitair Medische Centra. (2004a). *Onderzoek onderzocht: een bibliometrische analyse van het onderzoek van de universitaire medische centra*. Utrecht: Nederlandse Federatie van Universitair Medische Centra.

Nederlandse Federatie van Universitair Medische Centra. (2004b). *Van vele markten thuis. De Universitair Medische Centra in Nederland. Wat zijn ze, wat doen ze, wat willen ze*. Utrecht: Nederlandse Federatie van Universitair Medische Centra.

Noyons, E. C. M. (2001). Bibliometric mapping of science in a policy context. *Scientometrics, 50*(1), 83-98.

Noyons, E. C. M., Buter, R. K., & Van Raan, A. F. J. (1999). *Mapping the Field of Neurosciences* (No. TSER CT96-1035 (Project 1053), available via http://www.cwts.nl/ed/projects/home.htm). Leiden: Center for Science and Technology Studies.

Noyons, E. C. M., & Moed, H. F. (1999). Combining mapping and citations analysis for evaluative bibliometric purposes: a bibliometric study. *Journal of the American Society for Information Science, 50*(2), 115-132.

NRLO. (1997). *Een kwantitatieve analyse van de bodembiologie in Nederland*. Den Haag: Nationale Raad voor Landbouwkundig Onderzoek.

Nunnually, J. C. (1980). *Psychometric theory* (Vol. 2 nd edition): New York: McGraw-Hill.

NWO. (2001). *Themes plus talent: strategic plan 2002-2005*. Den Haag: NWO.

NWO Medische Wetenschappen, & Commissie Geneeskunde KNAW (1993). *De vormgeving van het Nederlands gezondheidsonderzoek*. Den Haag/Amsterdam.

NWO-MW. (1998). *Procedure-set*. Den Haag: NWO Medische Wetenschappen.

NWO-MW. (2000). *Experiment beloning referenten* (No. 0501exp.). Den Haag: NWO Medische Wetenschappen.

Nylenna, M., Andersen, D., Dahlquist, G., Sarvas, M., & Aakvaag, A. (1999). Handling of scientific dishonesty in the Nordic countries. *The Lancet, 354*(9172), 57-61.

Obbink, H. (2000). Bureaucratie rondom onderzoek smoort creativiteit. *Ad Valvas, Vrije Universiteit*.

OECD. (1992). *Autonomies and Interdependence in Science Policies. The changing role of Research Councils.* Paris: Group on scientific and university research. Meetings of heads of research councils of OECD Countries.

OECD. (1998). *Technology, Productivity and Job Creation: Best Policy Practices*. Paris:OECD.

OECD. (1999). *Changing structure, Organisation and nature of European PSR systems: European comparison of public research systems* (No. TSER Project No.SOE1-CT-96-1036). Paris: OECD.

Omta, S. W. F. (1995). *Critical success factors in biomedical research and pharmaceutical innovation.* Rijks Universiteit Groningen, Groningen.

Omta, S. W. F., & De Leeuw, A. C. J. (1997). Management control, uncertainty, and performance in biomedical research in universities, institutes and companies. *Journal of Engineering Technology Management, 14*(3-4), 223-257.

Overleg Manifest. (2002). Conclusies van het voorzitterschap. Barcelona.

Paolillo, J. G., & Brown, W. B. (1979). A multivariate approach to perceived innovation in R&D subsystems. *IEEE Transactions on Engineering Management, 26*(2), 36-39.

Patrick, J. P., & Stanley, E. C. (1996). Assessment of research quality. *Research in Higher Education, 37*(1), 23-42.

Pelz, D. C. (1956). Some social factors related to the performance of in a research organization. *Administrative Science Quarterly, 1*(2), 310-325.

Pelz, D. C., & Andrews, F. M. (1966). *Scientists in organizations: productive climates for research and development*. New York: John Wiley and Sons.

Perkoff, G. T. (1986). The research environment in family practice. *The journal of family practice, 21*(5), 389-393.

Peters, H. P. F., & Van Raan, A. F. J. (1993a). Co-word based science maps of chemical engineering, by combining clustering and multidimensional scaling. *Research Policy, 22*, 47-71.

Peters, H. P. F., & Van Raan, A. F. J. (1993b). Co-word based science maps of chemical engineering. Part 1: Representations by direct multidimensional scaling. *Research Policy, 22*(1), 23-45.

Pineau, C., & Levy-Leboyer, C. (1983). Managerial and organizational determinants of efficiency in biomedical research teams. In S. R. Epton, R. L. Payne & A. W. Pearson (Eds.), *Managing interdisciplinary research* (pp. 141-163). New York: John Wiley and Sons.

Pruthi, S., Jain, A., Wahid, A., & Nabi, S. A. (1993). Problems of management of scientific research: results of a survey. *Journal of Scientific & Industrial Research, 52*, 81-94.

Raad voor gezondheidsonderzoek. (2000a). *Activiteiten Raad voor gezondheidsonderzoek 1987-2000*. Den Haag: Raad voor gezondheidsonderzoek.

Raad voor gezondheidsonderzoek. (2000b). *Werkprogramma Raad voor gezondheidsonderzoek 2000-2001*. Den Haag: Raad voor gezondheidsonderzoek.

Raad voor gezondheidsonderzoek. (unknown). *Samenstelling Raad voor gezondheidsonderzoek*. Den Haag: Raad voor gezondheidsonderzoek.

Rennie, D. (1999). *Editorial peer review: its developments and rationale*. London: BMJ Books.

Reskin, B. F. (1979). Academic sponsorship and scientists' careers. *Sociology of Education, 52*(2), 129-146.

Rigter, H. (1986). Evaluation of performance of health research in the Netherlands. *Research Policy, 15*(1), 33-48.

Rip, A. (1987). Strategische sturing van wetenschap: noodzaak en (on)mogelijkheden. In R. J. Bijleveld & F. A. Van Vught (Eds.), *Sturing van hoger onderwijs* (39-59). 's-Gravenhage: VUGA Uitgeverij B.V.

Rip, A. (1990). Societal construction of research and technology. In H. Krupp (Ed.), *Technikpolitik angesichts der Umweltkatastrophe* (pp. 84-94). Heidelberg: Physica-Verlag.

Rip, A. (1999a). Fashions, lock-ins and the heterogeneity of knowledge production. *Futures*.

Rip, A. (1999b). *Research management from the top and from below* (Module 406 for M.Phil (Science and Technology Studies). Enschede: University of Twente.

Rip, A., & Van der Meulen, B. J. R. (1995). The patchwork of the Dutch evaluation system. *Research Evaluation, 5*(1), 45-53.

Rip, A. J. (1988). Mapping of science: possibilities and limitations. In A. F. J. Van Raan (Ed.), *Handbook of quantitative studies of Science and Technology* (pp.253-273). Amsterdam: Elsevier Science Publishers.

Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience: is agreement between reviewers any greater than would be expected by chance alone? *Brain, 123*(9), 1964-1969.

Ryan, J. C. (2003). The contribtution of cognitive psychology and organisational psychology to our understanding of scientific performance. In J. H. Hurley (Ed.), *Scientific research effectiveness: the organizational dimension* (pp. 197-213). Dordrecht/Boston/London: Kluwer Academic Publishers.

Saxberg, B. O., & Newell, W. T. (1983). Interdisciplinary research in the university: need for manegerial leadership. In S. R. Epton, R. L. Payne & A. W. Pearson (Eds.), *Managing interdisciplinary research* (pp. 202-210). New York: John Wiley and Sons.

Shinn, T. (2002). The triple helix and new production of knowledge: Prepackaged thinking on science and technology. *Social Studies of Science, 32*(4), 599-614.

Shrum, W., I. Chompalov and Genuth. (2001). Trust, Conflict and Performance in Scientific Collaborations. *Social Studies of Science 31*(5), 681-730.

Sinderman, C. J. (1985). *The joy of science; excellence and its rewards*. New York: Plenum Press.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between publications. *Journal of the American Society for Information Science, 24*, 265-269.

Smith, C. (1971). Age of R&D groups: a human consideration. *Human Relations, 23*, 81-96.

Smith, L. C. (1981). Citation analysis. *Library trends*, summer 1981, 83-106.

Smith, R. (1998). Unscientific practice flourishes in science. *British Medical Journal, 316*, 1036.

Smith, R., & Fiedler, F. E. (1971). The measurement of scholarly work: a critical review of the literature. *Educational record, 52*, 225-232.

Sonnert, G. (1995). What makes a good scientist?: determinants of peer evaluations among biologists. *Social Studies of Science, 25*, 35-55.

Spaapen, J. (1995). *Research for Society: Towards a method for the assessment of research in a policy context.* University of Amsterdam, Amsterdam.

Stankiewicz, R. (1976a). *Resarch groups and the academic research organization.* Paper presented at the 13th workshop of the international comparative study on the organization and performance of research units, Vienna.

Stankiewicz, R. (1979). The size and age of Swedish academic research groups and their scientific performance. In F. M. Andrews (Ed.), *Scientific productivity: the effectiveness of research groups in six countries* (pp. 191-222). Cambridge: Cambridge University Press/UNESCO.

Stankiewicz, R. (1980). *Leadership & the performance of research groups*. Research Policy Insitute University of Lund, Lund.

Stolte-Heiskanen, V. (1979). External resources and effectiveness. In F. Andrews (Ed.), *Scientific productivity: the effectiveness of research groups in six different countries* (pp. 121-153). Cambridge: Cambridge University Press/UNESCO.

Thamhain, H. J., & Wilemon, D. C. (1987). Building high performing engineering project teams. *IEEE Engineering Management Review, 34*(3), 130-137.

Unknown. (2002). *Onderzoek Nederland, 80*, 1-2.

Ugolini, D., Parodi, S., & Sant, L. (1997). Analysis of publication quality in a cancer research institute. *Scientometrics, 38*(2), 265-274.

van de Meent, Y., & van Vliet, R. (2005, 5-6 februari). Financieel prikkelen van onderzoekers is riskant: langs de lat. *NRC*.

van der Meulen, B. (1992). *Beoordelingsprocessen in wetenschap: de constructie van kwaliteit door wetenschap, industrie en overheid*. Universiteit Twente, Enschede.

Van der Meulen, B., & Rip, A. (1995). *Indicatoren en indicaties voor de beoordeling van maatschappelijke kwaliteit van onderzoek*. Centrum voor Studies van Wetenschap, Technologie en Samenleving, Universiteit Twente, Enschede.

Van der Meulen, B., & Rip, A. (2000). Evaluation of societal quality of public sector research in the Netherlands. *Research Evaluation, 8*(1), 11-25.

Van der Meulen, B. J. R., & Rip, A. (1994). *Research institutes in transition*. Centre for Studies of Science, Technology and Society, University of Twente, Enschede.

Van der Weijden, I. C. M. (1999a). *Kwaliteitsbeoordeling van (wetenschappelijk) gezondheidsonderzoek: opzet van een promotie onderzoek*. Vrije Universiteit, Faculteit Exacte Wetenschappen, Amsterdam.

Van der Weijden, I. C. M. (1999b). *Quality Assessment Systems in Scientific Medical and Pharmaceutical Research*. Paper presented at the 9-th European Summer School in Technology Management, University of Twente, Enschede.

Van der Weijden, I. C. M., Groenewegen, P., & Klasen, E. C. (2001). *Intermediaire organisaties in het Nederlands gezondheidsonderzoek: kwaliteitsbeoordeling van gezondheidsonderzoek*. Vrije Universiteit, Faculteit Exacte Wetenschappen, Amsterdam.

Van der Weijden, I. C. M., Groenewegen, P., & Klasen, E. C. (2002). Van beoordelingslast naar beoordelingslust? Vermindering belasting voor gezondheidsonderzoekers door externe beoordelingsystemen. *Tijdschrift voor Wetenschap, Technologie en Samenleving, 5*(3), 121-128.

Van der Weijden, I. C. M., Groenewegen, P., Van Ark, G., Ellenbroek, S. P. H., Smeenk, J.-W., & Stukart, M. (2002). *Small research council-Large added value? Quality assessment in medical and health research in the Netherlands.* Paper presented at the 5-th conference of the European Sociology Association, Helsinki.

Van Erp Taalman Kip, E. B. (1993). *Kwaliteitsystemen voor het laboratorium.* Paper presented at the R&D management conference, Amsterdam.

Van Leeuwen, T. N. (1996). *Bibliometric profiles of academic physics research in the Netherlands*. Centre for Science and Technology Studies, University of Leiden.

Van Raan, A. F. J. (1985). *Kwaliteit in wetenschap: een meetbaar begrip?* Leiden: DSWO Press.

Van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36,* 397-420.

Van Raan, A. F. J. (1999a). Evaluation of performance and trends in basic and applied research by advanced bibliometric methods: a science policy instrument for nations with economy in transition. In: V. Paes, L. Pivec, and A.H. Teich (eds.), *Science Evaluation and Its Management.* NATO Science Series 4: Science and Technology Policy, Vol.28. Amsterdam, IOS Press, 1999, pp. 227-245.

Van Raan, A. F. J. (1999b). *Evaluatie van wetenschappelijk onderzoek: beoordeling door vakgenoten versus bibliometrische technieken, invloed op onderzoeksmanagement en wetenschapsbeleid.* Paper presented at the conference ' Ontwikkelingen van wetenschappelijke en technologische kennis in het spanningsveld van de fin-de-siecle maatschappij' , Universiteit Leiden, Leiden.

Van Raan, A. F. J. (1999c). Bibliometrische methoden bij evaluatie en observatie van wetenschappelijke ontwikkelingen. In: G.M. van Trier, D.W.K. Jansen, H. Prins (eds), *Handboek Informatiewetenschap,* Samsom, Alphen aan den Rijn, I 700, pp. 1-36.

Van Raan, A. F. J., & Tijssen, R. J. W. (1993). The neural net of neural network research: an exercise in bibliometric mapping. *Scientometrics, 26*(1), 169-192.

Van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effects of open peer review on the quality of reviewers and on reviewers' recommendations: a randomised trail. *British Medical Journal, 318,* 23-27.

Van Rossum, W. (1987). *Sturing van wetenschap: de rol van onderzoeksorganisaties*. 's-Gravenhage: Staatsdrukkerij.

Van Steen, J. (2000). *S&T indicators in Dutch policy: a changing perspective.* Paper presented at the 6th International Conference on S&T Indicators, Leiden.

Vermeulen, M. (2002). Researchcode in het Academisch Medisch Centrum te Amsterdam. *Nederlands Tijdschrift voor Geneeskunde, 146*(35), 1620-1622.

Visart, N. (1979). Communication between and within research units. In F. Andrews (Ed.), *Scientific productivity: the effectiveness of research groups in six countries* (pp. 223-252). Cambridge: Cambridge University Press.

Vroeijensteijn, A. I. (1995). *Improvement and accountability: navigating between scylla and charybdis, Guide for external quality assessment in higher education*. London: Jessica Kingsley Publishers.

VSNU. (1996). *Afstemming onderzoeksbeoordelingen: rapportage tripartite werkgroep afstemming onderzoeksbeoordelingen*. Vereniging van Universiteiten, Utrecht.

VSNU, NWO, & KNAW. (2003). *Standard Evaluation Protocol 2003-2009 for Public Research Organisations*. VSNU/NWO/KNAW, Utrecht/Den Haag/Amsterdam.

Wagner, C. (1999). *Implementation and effectiveness of quality systems in Dutch health care organizations*. Vrije Universiteit, Amsterdam.

Wallmark, J. T., Eckerstein, S., LAngered, B., & Holmqvist, E. S. (1973). The increase in efficiency with size of research teams. *IEEE Transactions on Engineering Management, 20*(3), 80-86.

Wallmark, J. T., & Sellerberg, B. (1966). Efficiency vs size of research teams. *IEEE Transactions on Engineering Management, EM-13*(3), 137-142.

Wamelink, F., & Spaapen, J. (1999). *De evaluatie van universitair onderzoel: methodiek voor het incorporeren van de maatschappelijke waarde van onderzoek* (No. NRLO 99/12). Nationale Raad voor Landbouwkundig Onderzoek, Den Haag.

Wells, W. P. (1962). *Group age and scientific performance*. University of Michigan, Michigan.

Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature, 387*, 341-343.

Werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek. (2000). *Kwaliteit verplicht: naar een nieuw stelsel van kwaliteitszorg voor wetenschappelijk onderzoek*. KNAW, Amsterdam.

Wessely, S. (1998). Peer review of grant applications: what do we know? *The Lancet, 352*, 301-305.

West, M. A., Smith, H., Feng, W. L., & Lawthom, R. (1998). Research excellence and departmental climate in Brittish universities. *Journal of Occupational and Organizational Psychology, 71*, 261-281.

Westerheijden, D. F. (1997). A solid basis for decisions. *Higher Education, 33*, 397-413.

Wood, F. Q. (1997). *The peer review process* (No. 54). Canberra: National board of employment, education and training.

Wouters, P. (1999). *The Citation Culture.* University of Amsterdam, Amsterdam.

Zeldenrust, S. (1988). *Ambiguity, choice and control in research*. University of Amsterdam, Amsterdam.

ZON. (1999). *Procedures voor ZON-programma's*. ZorgOnderzoek Nederland, Den Haag.

ZonMw. (2002). *Procedures*. ZonMw, Den Haag.

ANNEX 1

Interview question list

Depending on the interview partners, different questions were asked and different topics were discussed. Interviews with representatives of intermediary organisations and research administrators generally consisted of three parts: a first introductory part, a second part about the research evaluations used and a third part dealing with the effects of the evaluations. Interviews with research leaders generally consisted of five parts: an introductory part, a part about research funding, an explorative part about research management issues and, finally, a fourth and a fifth part dealing with internally and externally organised research evaluations respectively.

I. Representatives of intermediary organisations

General questions:

- What are the goals and functions of the intermediary organisation?
- What is the area of attention of the intermediary organisation?
- When was the intermediary organisation established?
- Why was the intermediary organisation established?
- What is your current position within the intermediary organisation?
- How long have you been working for the intermediary organisation?
- How is the concept 'research quality' defined within the intermediary organisation?

Research evaluations

Present research evaluations:

- In which phases of the research process (input, throughput and/or output) has the health research been evaluated?
- What (kind of) procedures were used in the evaluation of health research?
- What criteria and indicators were used in the evaluation of health research?
- What are the most important indicators or criteria in the evaluation of health research?
- Do the evaluation procedures used distinguish scientific quality from societal quality? If so, does the intermediary organisation use separate procedures, protocols and indicators?
- In evaluating research, to what extent does the intermediary organisation take the outcomes and results of previous research evaluations into account?

Changes and developments in research evaluations:

- What changes or developments have recently taken place in research evaluation procedures (for example, changes in criteria, indicators, frequency, evaluation unit)?
- What are the reasons for these changes or developments?
- What is the opinion of the intermediary organisation about the developments that have recently taken place in the evaluation of health care research?
- What changes or developments in evaluation procedures of health research will take place in the near future?

Impact

- What, from the point of view of the intermediary organisation, are the implications of the procedures used in evaluating health research for health research groups in the Netherlands?
- What, from your point of view of the intermediary organisation, are the implications of the results and outcomes of health research evaluations for health research groups in the Netherlands?

II. Research administrators

General questions:

- What is your current position within the organisation?
- How long have you been working for the organisation?
- How is the faculty or research organisation structured?
- How is the research budget allocated to the research groups?
- How is the concept 'research quality' defined within the organisation?

Research evaluations

Present research evaluations:

- In which phases of the research process (input, throughput and/or output) is health research evaluated?
- What procedures and protocols are used in the evaluation of health research?
- What criteria and indicators are used in the evaluation of health research?
- What is the most important indicator or criterion in the evaluation of health research?
- Do the evaluation procedures used distinguish scientific quality from societal quality? If so, does the faculty or research organisation use separate procedures, protocols and/or indicators?

Changes and developments in research evaluations:

- What changes or developments in the evaluation procedures of health research will take place in the near future?
- What are the reasons for these changes or developments?

Impact of internally organised research evaluations:

- How did the faculty or research organisation respond to the outcomes of the Discipline Report on medical and Health Sciences Research in the Netherlands 1998 (DAG 1998)?
- What decisions have been made as a result of the DAG 1998?
- How important were other external research evaluations such as evaluations of research schools, evaluation of NWO and/or ZON research proposals and bibliometric analysis in this decision-making process?
- How could research groups themselves implement the results and recommendations of the DAG 1998?

Impact

- In what way(s) does the faculty or research organisation communicate with research groups?
- Has the faculty or research organisation some direct or indirect influence on the direction and quality of health research (conducted within research groups)?

Future research

- Are you willing to assist me in preparing future research by collecting the names and addresses of research groups?

III. Research leaders

General questions

- Which type of medical and health research (fundamental, strategic and applied) is conducted within the research group?
- In which different medical disciplines can your research group be placed?
- What is the size (number of employees in fte) of your research group?
- How many research staff are employed relative to the number of administrative staff?
- How could you formulate the mission of your research group?
- Could you name the two most important scientific journals you read to keep informed about recent developments?

- Could you name the two most important scientific journals in which your research group publish articles?
- Could you mention the most important national and international research meeting (e.g. conference, symposium) in your research area?

Research funding

- From which of the following external research funding agencies does your research group obtain funding: NWO (including ZON), KNAW, charity funds, firms and industries, European funds, other external funding agencies?
- What criteria are used in selecting external funding agencies?
- Have opportunities for receiving external research funding recently changed (in the case of your research group)?

Research management

- Could you indicate whether research management issues given below could influence research quality?
  - staffing policy
  - collaborations
  - scientific strategy
  - consistency of research topics
  - flow of PhD students and other scientific staff to other research groups
  - participation in a research school

Internally organised research evaluations

- Do you evaluate research staff (e.g. job assessments)?
- Are research proposals pre-assessed in your research group before submission to external funding agencies?
  - What procedures and protocols are used in pre-evaluation of research proposals?
  - What criteria and indicators are used in pre-evaluation of research proposals?
- Is research output evaluated in your research group?
  - What procedures and protocols are used in the pre-evaluation of research proposals?
  - What criteria and indicators are used in pre-pre-evaluation of research proposals?
- In what other internally organised research evaluations does your group participate?
  - Research output evaluations organised by the medical faculty

– Research output evaluations organised by the university or non-university research institute

– Pre-evaluations of research proposals organised by the medical faculty

– Pre-evaluations of research proposals organised by the university or non-university research institute

- How, in general, are the results and recommendations of internally organised research evaluations implemented?

Externally organised research evaluations

- In what other externally organised research evaluations does your group participate?
  – Discipline Report on medical and Health Sciences Research in the Netherlands 1998
  – Accreditation of a research school
  – Research output evaluation of research schools
- How, in general, are the results and recommendations of these externally organised research evaluations implemented?

Views on the importance of research evaluations

- Which research evaluation is the most important?
- Which research evaluation is the least important?
- Has the realisation that ‘research quality is of vital importance’ increased, decreased or not changed during the past few years?

ANNEX 2

Abbreviations

AHCI	Arts and Humanities Citation Index
ALIFI	Amsterdam-Leiden Institute For Immunology
AMC	Amsterdam Medical Centre
ANOVA	One-way Analysis of Variance
ECOS	Commission for the Recognition of Research Schools
EMC	Erasmus Medical Centre
CBF	Dutch Central Bureau for Fund-raising
CDAG	Dutch Committee for Disciplinary Research
CI	Citation Indices
COS	Cooperating Sector Council
CWTS	Centre for Science and Technology Studies
DMW-VSNU	Association of universities in the Netherlands, Council for medical faculties in the Netherlands
DAG	Discipline Report on medical and Health Sciences Research
ECOS	Commission for the Recognition of Research Schools
EU	European Union
FTE	Full-Time Equivalent
FMWV	Federation of Biomedical Scientific Societies
GCP	Good Clinical Practice
GDP	Gross Domestic Product
GLP	Good Laboratory Practice
GMC	Groningen Medical Centre
GMP	Good Manufacturing Practice
KNAW	Royal Netherlands Academy of Arts and Sciences
KWO	Workgroup on Quality Assurance of Scientific Research
IAC	International Assessment Committee
IF	Impact Factor
ISI	Institute for Scientific Information
LUMC	Leiden University Medical Center
MSCI	Material Science Citation Index

NCI	Neuroscience Citation Index
NFU	Netherlands Federation of University Medical Centers
NIVEL	Netherlands Institute for Health Services Research
NWO	Netherlands Organization for Scientific Research
MW-NWO	Department of Medical Sciences of the Netherlands Organisation for Scientific Research
NOD	Dutch Research Data base
OECD	Organization for Economic Co-operation and Development
PCA	Principal Component Analysis
PIONIER	Personal Impulse for Research Groups with Ideas for Excellent Research
QANU	Quality Assurance for Netherlands Universities
R&D	Research and Development
RAE	Research Assessment Exercise
RB	Research Board
REPP	Research Embedment and Performance Profile
RGO	Advisory Council on Health Research
RIVM	National Institute of Public Health and the Environment
RMW	Council for Medical Sciences (Royal Netherlands Academy of Arts and Sciences)
SCI	Science Citation Index
SQRP	Societal Quality Research Profile
SSCI	Social Science Citation Index
TNO	Netherlands Organization for Applied Scientific Research
TQM	Total Quality Management
UK	United Kingdom
UMC	Utrecht Medical Centre
UMCN	University Medical Centre Nijmegen
US	United States
VFI	Dutch Association of Fund-raising Organizations
VSNU	Association of Universities in the Netherlands
VUMC	Vrije Universiteit Medical Centre
ZON	Netherlands Organisation for Health Care Research and Development Council
ZonMw	Netherlands Organisation for Health Research and Development

ANNEX 3

List of persons interviewed

Ark, G. van	NWO-MW	18-04-2000
Benneker, H.	Advisory Council on Health Research	28-03-2000
Bickerstaffe, R.	Solvay Pharmaceuticals	06-05-1999
Collijn, D.	Maastricht University	01-11-2000
Cools, A.	University of Nijmegen	29-01-2001
Coppens, M.	VSNU	09-03-2000
Daha, M.R.	Research School ALIFI	26-02-1999
Deen, K.	KNAW Council for Medical Sciences	11-05-2000
Dijk, F. van	University of Amsterdam	22-03-2001
Egberink, G.	University of Nijmegen	14-02-2001
Festen, C.	Erasmus University Rotterdam	27-10-2000
Groenewegen, P.P.	NIVEL	14-03-2001
Guttinger, V.	TNO, Prevention and Health	15-05-2001
Haan, P. de	Vrije Universiteit Amsterdam	13-03-2001
Havekes, L.	TNO, Prevention and Health	19-03-2001
Joling, W.	ZON	12-04-2000
Klasen, E.C.	NWO-MW	23-04-1999/11-05-1999
Kukenheim, R.	Leiden University	03-11-2000
Kraal, G.	Research School ALIFI	02-03-1999
Meulen, B.J.R. van der	University of Twente	08-04-1999
Miedema, F.	Sanquin	24-02-1999
Neefjes, J.	Netherlands Cancer Institute	26-01-2001
Nicholson, N.V.	Organon	28-04-1999
Nieboer, H.	TNO Nutrition and Food Research	29-04-1999
Oostra, B.	Erasmus University Rotterdam	12-01-2001
Plasterk, R.	Netherlands Cancer Institute	09-06-1999
Poppema, S.	University of Groningen	23-04-2001
Raan, A.J.R. van	Leiden University	06-04-1999
Roseboom, H.	Solvay Pharmaceuticals	06-05-1999
Ruitenberg, E.C	Sanquin	08-03-1999/11-08-1999
Sanderman, R.	University of Groningen	23-04-2001
Schlotter, R.	Utrecht University	06-04-2001
Schuitemaker, H.	Sanquin	22-10-2000

Smeenk, J.W.	ECOS/ Sanquin	15-02-2000/21-03-2001
van Steijn, F. van	VSNU	09-03-2000
Stukart, M.	KNAW Council for Medical Sciences	11-02-2000
Tanke, H.	Leiden University	28-06-2001
Thiens, T.	University of Nijmegen	14-02-2001
Timmerman, H.	Vrije Universiteit Amsterdam	04-03-1999
Visschedijk-Brinkman, M.	TNO Nutrition and Food Research	29-04-1999
Visser, C.	Vrije Universiteit Amsterdam	17-05-2001
Vos, E.	Netherlands Cancer Institute	23-10-2000
Voest, E.	Utrecht University	21-03-2001
Wagstaff, J.	Maastricht University	19-01-2001
Zegers, B.	Research School Infection and Immunity	03-03-1999

ANNEX 4

Questionnaire

SECTION I: RESPONDENT INFORMATION

Please indicate your sex: 1 ❑ Male
2 ❑ Female

Question 1. What is your year of birth? 19—

Question 2a. In what year did you obtain your PhD? 19—

Question 2b. How many years of experience do you have in health research? ---years

Question 2c. How many years of foreign health research experience do you have? ---years

Question 2d. Did you ever hold a management position?
1 ❑ No
2 ❑ Yes ---years

Question 3. Please indicate your workplace:
1 ❑ University
2 ❑ Non-university research institute

Question 4a. How frequently does the composition of your research group change in time?
1 ❑ The composition of the research group has been stable for several years
2 ❑ The composition of the research group has been stable for at least one year
3 ❑ The composition of research groups changes every 6 months
4 ❑ The composition of the research group changes frequently

Question 4b. Does your research group conduct more than one research project at the same time?
1 ❑ Yes
2 ❑ No

Question 4c: Are you the only research leader of your research group?
1 ❑ Yes
2 ❑ No, I share the leadership with others

Question 4d: How long have you been directing this research group? ---years

SECTION II: GENERAL QUESTIONS FOR THE RESEARCH GROUP

Question 5a. In which of these areas can your research group be placed? **(Please tick all suitable answers)**

1 ❑ Cell biology and developmental biology
2 ❑ Genetics
3 ❑ Bio-informatics and epidemiology
4 ❑ Immunology
5 ❑ Microbiology
6 ❑ Virology
7 ❑ Oncogenesis
8 ❑ Cardiovascular system
9 ❑ Haematology
10 ❑ Nephrology
11 ❑ Respiration
12 ❑ Dermatology
13 ❑ Musculoskeletal disorders
14 ❑ Gastroenterology and hepatology
15 ❑ medical technology
16 ❑ Pharmacology and toxicology
17 ❑ Endocrinology
18 ❑ Metabolism
19 ❑ Basal neurology
20 ❑ Neurology
21 ❑ Psychiatry
22 ❑ Medical psychology
23 ❑ Environment, work and health
24 ❑ Gerontology and geriatrics
25 ❑ Youth and health
26 ❑ Social medicine
27 ❑ General practice
28 ❑ Public health

Question 5b. Which of the areas mentioned in question 5a is most important in characterising your research group? *(Please tick the most appropriate answer)* ----------

Question 5c. On which part of health research does your research group mainly concentrate? ***(Maximum of two answers are allowed)***

1 ❑ Subcellular level
2 ❑ Cell level
3 ❑ Organ level
4 ❑ Pathological process
5 ❑ Patients
6 ❑ Public health/health care system

Question 5d. Please describe the proportion of research conducted by your research group that you characterise as multidisciplinary.

1 ❑ 0–20%
2 ❑ 21–40%
3 ❑ 41–60%
4 ❑ 61–80%
5 ❑ 81–100%

Question 6a. Please indicate below the number of employees (fte) in your research group (including yourself).

	PERMANENT APPOINTMENT	**TEMPORARY APPOINTMENT**
Professors		
Senior scientific staff		
PhD students		
Laboratory technical staff		
Technical support staff		
Other scientific staff		

Question 6b. How many staff are working under your leadership in the research group? --- fte

Question 7a. Please indicate the three most important aims of your research group.

1 ❑ Scientific publications
2 ❑ Exploitation of knowledge
3 ❑ Development of methodology
4 ❑ Training of young researchers
5 ❑ Development of research abilities
6 ❑ Development of new knowledge
7 ❑ Development of new products
8 ❑ New collegiate interaction
9 ❑ Knowledge transfer to users
10 ❑ Collection of empirical material
11 ❑ Gaining international contacts
12 ❑ Receiving research funding

Question 7b. What is your attitude towards the following statements about the outcomes of research conducted in your research group?

	Never	**Rarely**	**Sometimes**	**Often**	**Always**
Research results in Dutch publications	❑	❑	❑	❑	❑
Research results in international publications	❑	❑	❑	❑	❑
Research results in clinical medical applications (e.g. development and improvement of medicine or treatment)	❑	❑	❑	❑	❑
Research results in general medical applications	❑	❑	❑	❑	❑

Question 8. Please indicate (by circling a figure) your attitude towards the following statements about choosing new research topics

	Not Important		**Neutral**		**Very Important**
Connection with my previous research	1	2	3	4	5
Continuation of research themes in my discipline	1	2	3	4	5
Connection with internationally important research themes	1	2	3	4	5
Theoretically challenged and innovative research themes	1	2	3	4	5
Possibility of testing totally new research themes	1	2	3	4	5
Interest and possibilities for young researchers	1	2	3	4	5
Practicability of research by young researchers	1	2	3	4	5
The guarantee that PhDs will be obtained	1	2	3	4	5
Opinion of colleagues in the Netherlands	1	2	3	4	5
Opinion of foreign colleagues	1	2	3	4	5
Spearheads of Dutch research programmes	1	2	3	4	5
Spearheads of European research programmes	1	2	3	4	5
Possibility of applying research to Dutch health care	1	2	3	4	5
Possibility of obtaining visibility in top journals (e.g. Nature, Science, Lancet)	1	2	3	4	5
Use of available or new lab equipment	1	2	3	4	5

Question 9. Please indicate whether research conducted by your research group is part of the core business of your faculty or research institute.

1 ❑ Yes
2 ❑ No
3 ❑ Do not know

Question 10a. Please indicate below how much time your research group spends on average on:

Research	❑ 1–20%	❑ 21–40%	❑ 41–60%	❑ 61–80%	❑ 81–100%
Education	❑ 1–20%	❑ 21–40%	❑ 41–60%	❑ 61–80%	❑ 81–100%
Patient care	❑ 1–20%	❑ 21–40%	❑ 41–60%	❑ 61–80%	❑ 81–100%

Question 10b. Please indicate below how much time you spend on average on:

	None	1–10%	11–20%	21–30%	31–40%	41–50%	51–100%
Laboratory experiments and analyses	❑	❑	❑	❑	❑	❑	❑
Education	❑	❑	❑	❑	❑	❑	❑
Patient care	❑	❑	❑	❑	❑	❑	❑
Supervision (PhD students)	❑	❑	❑	❑	❑	❑	❑
Internal research management	❑	❑	❑	❑	❑	❑	❑
External research management	❑	❑	❑	❑	❑	❑	❑
Other research activities	❑	❑	❑	❑	❑	❑	❑

Question 11. What is your opinion on the following general statements about research involvement?

	Dis-Agree fully	Dis-agree partly	Neither agree nor disagree	Agree partly	Agree fully
1. I feel more like a researcher than like a manager	❑	❑	❑	❑	❑
2. I feel still involved in research conducted by my research group	❑	❑	❑	❑	❑
3. I'm still well informed about recent developments taking place in my research area	❑	❑	❑	❑	❑
4. I keep up with the specialist literature	❑	❑	❑	❑	❑
5. By generating new ideas I still contribute to research conducted by my research group	❑	❑	❑	❑	❑
6. I'm intensively involved in more research project(s) conducted by my research group	❑	❑	❑	❑	❑
7. I regularly attend research meetings organised in my research group	❑	❑	❑	❑	❑
8. I regularly publish as first author in international journals	❑	❑	❑	❑	❑
9. My staff think of me as a highly skilled scientist	❑	❑	❑	❑	❑
10. I'm acting as oracle in solving research problems	❑	❑	❑	❑	❑

Question 12. Please indicate below the most important articles your research group has published in the past five years

Journal	----------	Journal	----------	Journal	----------
Year	----------	Year	----------	Year	----------

SECTION III: RESEARCH FUNDING

Question 13. Please indicate the (average) yearly research budget of your research group (in Dutch guilders).

1 ❑ 1.000.000–2.000.000
2 ❑ 2.000.000–5.000.000
3 ❑ 5.000.000–10.000.000
4 ❑ more than 10.000.000

Question 14. Please indicate the proportion of your research group's funds that come from:

University	----------%	Ministries	----------%
Research institute	----------%	Charity funds	----------%
Research Council NWO	----------%	Firms/industry	----------%
Research Council ZON	----------%	European funds	----------%
Royal Netherlands Academy of Arts and Sciences	----------%	Others	----------%

Question 15. This question concerns your research group. Compared to five years ago:

		De-creased	Not changed	In-creased	Don't know
1.	The amount of money received from the first flow has	❑	❑	❑	❑
2.	The amount of money received from NWO has	❑	❑	❑	❑
3.	The number of submitted research proposals to NWO has	❑	❑	❑	❑
4.	The amount of money received from charity funds has	❑	❑	❑	❑
5.	The number of submitted research proposals to charity funds has	❑	❑	❑	❑
6.	The amount of money received from Dutch ministries has	❑	❑	❑	❑
7.	The number of submitted research proposals to EU funds has	❑	❑	❑	❑
8.	The amount of money received from EU funds has	❑	❑	❑	❑
9.	The amount of money received from firms, industries and health insurers has	❑	❑	❑	❑

Please think about the last time your research group submitted a research proposal to NWO (research questions 16-19).

Research question 16. Please indicate the result of the evaluation of the research proposal.

1 ❑ Rejected in pre-evaluation
2 ❑ Accepted in pre-evaluation, rejected in evaluation
3 ❑ Accepted
4 ❑ I have never submitted a research proposal to NWO

Research question 17[1]. To what extent were you informed about the evaluation procedures?
Please indicate your response by circling a figure

To a large extent 1-2-3-4-5 to a small extent

Research question 18[1]. The following items refer to the procedures used to evaluate a research proposal submitted. To what extent:

1. Have you been able to express your views and feelings during the procedure? To a large extent 1-2-3-4-5 to a small extent
2. Have you had influence on the evaluation outcome? To a large extent 1-2-3-4-5 to a small extent
3. Has the evaluation been applied consistently? To a large extent 1-2-3-4-5 to a small extent
4. Has the evaluation been free of bias? To a large extent 1-2-3-4-5 to a small extent
5. Has the evaluation been based on accurate information? To a large extent 1-2-3-4-5 to a small extent
6. Has the evaluation upheld ethical and moral standards? To a large extent 1-2-3-4-5 to a small extent

Research question 19[1]. The following items refer to the outcome of the evaluation of a research proposal. To what extent:

1. Does the outcome reflect the effort you have put into the proposal? To a large extent 1-2-3-4-5 to a small extent
2. Is the outcome appropriate in the light of the work you have completed? To a large extent 1-2-3-4-5 to a small extent
3. Does the outcome reflect your contribution to the research field? To a large extent 1-2-3-4-5 to a small extent
4. Is the outcome justified given your performance? To a large extent 1-2-3-4-5 to a small extent

SECTION IV: RESEARCH MANAGEMENT

Question 20. Please indicate whether you have had problems in securing qualified scientific staff.

	Never	Rarely	Sometimes	Often	Always	Not applicable
Professors	❑	❑	❑	❑	❑	❑
Senior scientific staff	❑	❑	❑	❑	❑	❑
PhD students	❑	❑	❑	❑	❑	❑
Laboratory technician staff	❑	❑	❑	❑	❑	❑
Technical support staff	❑	❑	❑	❑	❑	❑
Other scientific staff	❑	❑	❑	❑	❑	❑

Question 21a. What efforts does the research group make to secure qualified scientific staff?

1. ----------
2. ----------
3. ----------

Question 21b. Are the efforts mentioned in question 21a also made by your university or research institute?

1 ❑ Never
2 ❑ Rarely
3 ❑ Sometimes
4 ❑ Often
5 ❑ Always

Question 22a. Please indicate to what extent the research group offers rewards to their employees.

	Never	Sometimes	Often
To take (inter)national courses and education	❑	❑	❑
To attend (inter)national conferences	❑	❑	❑
To gain experience in foreign research groups	❑	❑	❑
To supervise master and PhD students	❑	❑	❑
To receive financial bonuses	❑	❑	❑
To work at home	❑	❑	❑
To work flexible hours	❑	❑	❑
To receive special commendations or prizes	❑	❑	❑

Question 22b. Are the efforts mentioned in question 22a also made by your university or research institute?

1 ❑ Never
2 ❑ Rarely
3 ❑ Sometimes
4 ❑ Often
5 ❑ Always

Question 23. Please indicate whether researchers are assessed by job assessments.

1 ❑ Yes
2 ❑ No

Question 24. How often are job assessments done in your research group?

1 ❑ Once every 6 months
2 ❑ Once a year
3 ❑ Once every 2 years
4 ❑ Once every 5 years

Question 25. What are the goals of job assessments in your research group?

1 ❑ To evaluate researcher output

2 ❑ To present possibilities for developing researcher skills
3 ❑ To make yearly appointments with the researcher
4 ❑ To reflect upon leadership received from the researcher
5 ❑ To develop a relationship between the researcher and the supervisor

Question 26. Please indicate the frequency of meetings held (within the research group) to discuss research?

	Never	**Once a year**	**Twice a year**	**Once a month**	**Once a week**
Discussions about literature	❑	❑	❑	❑	❑
Progress meetings about current research projects	❑	❑	❑	❑	❑
Research presentations	❑	❑	❑	❑	❑
Discussions about (conference) papers	❑	❑	❑	❑	❑
Discussions about research proposals	❑	❑	❑	❑	❑
Other research meetings…….	❑	❑	❑	❑	❑

Question 27. Please indicate what activities are organised in your research group *(Please tick all suitable answers)*

1 ❑ Lunches
2 ❑ Drinks
3 ❑ Retreats
4 ❑ Other activities……..

Question 28. To what extent do you use different means of communication?
1 ❑ E-mail
2 ❑ Telephone
3 ❑ Memorandum
4 ❑ Intranet
5 ❑ Personal, with appointment
6 ❑ Personal, without appointment

Question 29a. Please indicate whether your research group organises discussion meetings about research policy plans.
1 ❑ Yes
2 ❑ No, please go to question 29d

Question 29b. Which staff participate in discussions about research policy plans? *(Please tick all suitable answers)*
1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the same research school
4 ❑ Researchers employed in a different university or research institute
5 ❑ Foreign researchers

Question 29c. Why does your research group organise discussion about research policy plans?

Question 29d. Does the group's research policy plan form a part of the main points of the research policy plan formulated for the whole faculty or research institute?
1 ❑ Yes
2 ❑ No
3 ❑ Not applicable

Question 30. Please indicate whether your research group is facing problems/difficulties in conducting research due to: *(Please tick all suitable answers)*

1 ❑ Equipment
2 ❑ Research material
3 ❑ Data processing
4 ❑ Laboratory room
5 ❑ Library services
6 ❑ Personnel
7 ❑ Other infrastructure conditions----------

Question 31. How many working days (1 working day is equivalent to 8 hours) per year do you spend on average on external activities?

		None	**1–10 days**	**11–20 days**	**21–30 days**	**31–40 days**	**41–50 days**	**> 50 days**
1.	Working visits	❑	❑	❑	❑	❑	❑	❑
2.	Supervision of PhD students employed outside your research group	❑	❑	❑	❑	❑	❑	❑
3.	Presentation of lectures	❑	❑	❑	❑	❑	❑	❑
4.	Attending symposia and conferences	❑	❑	❑	❑	❑	❑	❑
5.	Organising symposia and conferences	❑	❑	❑	❑	❑	❑	❑
6.	Participation in editorial boards of scientific journals	❑	❑	❑	❑	❑	❑	❑
7.	Participation in assessment committees	❑	❑	❑	❑	❑	❑	❑
8.	Participation in audit committees	❑	❑	❑	❑	❑	❑	❑

Questions 32. Please indicate (with mark between 1, absolutely unimportant to 10, essential) how important collaboration is with:

1. Members of a research school Mark ----------
2. Dutch research groups Mark ----------
3. International research groups Mark ----------
4. Semi-government institutes Mark ----------
5. Pharmaceutical industries Mark ----------
6. Firms Mark ----------
7. Clinics Mark ----------
8. Family doctors Mark ----------
9. Ministries Mark ----------

Question 32b. What, from the point of view of your research group, are the three most important goals of scientific collaboration?

1 ❑ Scientific publications
2 ❑ Exploitation of knowledge
3 ❑ Development of methodology
4 ❑ Training of young researchers
5 ❑ Development of research abilities
6 ❑ Development of new knowledge
7 ❑ Development of new products
8 ❑ New collegiate interaction
9 ❑ Knowledge transfer to users
10 ❑ Collection of empirical material
11 ❑ Gaining international contacts
12 ❑ Receiving research funding

SECTION V: INTERNAL RESEARCH EVALUATIONS

Question 33a. Please indicate whether research proposals are pre-assessed internally before submission to external funding agencies.

1 ❑ No -> Please go to question 35b
2 ❑ Yes, at the level of my research group
3 ❑ Yes, at the level of the faculty or research institute (part of university)
3 ❑ Yes, at the level of the university or research institute

Question 33b. What research proposals are pre-assessed?

Question 34. Please indicate the judges of pre-assessments of research proposals. *(Please tick all suitable answers)*

Research group level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
4 ❑ Researchers employed in the Netherlands
5 ❑ Foreign researchers
6 ❑ Other judge(s)
7 ❑ Not applicable

Medical faculty level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the Netherlands
4 ❑ Foreign researchers
5 ❑ Other judge(s)
6 ❑ Not applicable

University or non-university research institute level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the Netherlands
4 ❑ Foreign researchers
5 ❑ Other judge(s)
6 ❑ Not applicable

Question 35. What is your opinion on the following statements about the pre-assessment of research proposals?

		Dis-Agree fully	**Dis-agree partly**	**Neither agree nor disagree**	**Agree partly**	**Agree fully**
1.	Researchers are not obliged to incorporate any comments and suggestions received from judges	❑	❑	❑	❑	❑
2.	Researcher leader decides whether research proposals are submitted	❑	❑	❑	❑	❑
3.	Faculty/research institute must have the opportunity to select and submit research proposals to external funding agencies	❑	❑	❑	❑	❑
4.	Internal pre-assessments of research proposals generally result in major changes to receiving external funding	❑	❑	❑	❑	❑

Question 36. Please indicate whether the research output of your group is evaluated internally. *(Please tick all suitable answers)*

1 ❑ Yes, at the level of my research group
2 ❑ Yes, at the level of the medical faculty
3 ❑ Yes, at the level of the university or non-university research institute
4 ❑ No -> please go to question 43

Question 37. Please indicate the judges of internally organised research output evaluation(s). *(Please tick all suitable answers)*

Research group level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the Netherlands
4 ❑ Foreign researchers
5 ❑ Other judge(s)
6 ❑ Not applicable

Medical faculty level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the Netherlands
4 ❑ Foreign researchers
5 ❑ Other judge(s)
6 ❑ Not applicable

University or non-university research institute level

1 ❑ Researchers employed in 'own' research group
2 ❑ Researchers employed in the same university or research institute
3 ❑ Researchers employed in the Netherlands
4 ❑ Foreign researchers
5 ❑ Other judge(s)
6 ❑ Not applicable

Question 38. Please indicate the frequency of internally organised research output evaluation(s). *(Please tick all suitable answers)*

Research group level

1 ❑ Once every 3–5 years
2 ❑ Once every 2–3 years
3 ❑ Once a year
4 ❑ Not applicable

Medical faculty level

1 ❑ Once every 3–5 years
2 ❑ Once every 2–3 years
3 ❑ Once a year
4 ❑ Not applicable

University or non-university research institute level

1 ❑ Once every 3–5 years
2 ❑ Once every 2–3 years
3 ❑ Once a year
4 ❑ Not applicable

Question 39. Please indicate which of the following criteria or indicators are used in internally organised research output evaluation(s). *(Please tick all suitable answers)*

Research group level

1 ❑ Number of publications
2 ❑ Impact factors of SCI journals
3 ❑ Citation analysis
4 ❑ Benchmark
5 ❑ External research funding received
6 ❑ Number of dissertations
7 ❑ Number of presentations
8 ❑ Not applicable

Medical faculty level

1 ❑ Number of publications
2 ❑ Impact factors of SCI journals
3 ❑ Citation analysis
4 ❑ Benchmark
5 ❑ External research funding received
6 ❑ Number of dissertations
7 ❑ Number of presentations
8 ❑ Not applicable

University or non-university research institute level

1 ❑ Number of publications
2 ❑ Impact factors of SCI journals
3 ❑ Citation analysis
4 ❑ Benchmark
5 ❑ External research funding received
6 ❑ Number of dissertations
7 ❑ Number of presentations
8 ❑ Not applicable

Question 40. Please indicate the goals of the internally organised research output evaluation(s).

Research group level

Medical faculty level

University or non-university research institute level

Question 41a. How are you experiencing the results and recommendations of internally organised research output evaluation(s) in general?

Very useful		**Neutral**		**Not Useful**
1	2	3	4	5

Question 41b. Please indicate whether the results and recommendations of internally organised research output evaluation(s) are incorporated?

Very intensively		**Neutral**		**Not intensively**
1	2	3	4	5

Question 42. Please indicate whether the incorporation of recommendations from internally organised research evaluations are tested in subsequent research evaluations.

1 ❑ Never
2 ❑ Rarely
3 ❑ Sometimes
4 ❑ Often
5 ❑ Always

SECTION V: EXTERNAL RESEARCH EVALUATIONS

Question 43. Please indicate whether your research group is involved in externally organised research output evaluations. *(Please tick all suitable answers)*

1 ❑ Discipline assessment of medical and health research in the Netherlands
2 ❑ Evaluations of a research school
3 ❑ Other externally organised research evaluation(s)

Question 44. Please indicate the result of your research theme evaluation in the national discipline assessment of medical and health research in 1998?

1 ❑ Poor (score 1)
2 ❑ Unsatisfactory (score 2)
3 ❑ Satisfactory (score 3)
4 ❑ Good (score 4)
5 ❑ Good to very good (score 5)
6 ❑ Very good (score 6)
7 ❑ Very good to excellent (score 7)
8 ❑ Excellent (score 8)
9 ❑ Not assessed in the national discipline assessment of medical and health research in 1998

Question 45. Please indicate to what extent the score of your research theme, in the national discipline assessment of medical and health research of 1998, corresponds to your own estimation?

1 ❑ Score is lower than my own estimation
2 ❑ Score corresponds to my own estimation
3 ❑ Score is higher than my own estimation
4 ❑ Don't know

Question 46. To what extent were you informed about the procedures used in the national discipline assessment of medical and health research of 1998? *Please indicate your response by circling a figure*

To a large extent 1-2-3-4-5 to a small extent

Question 47a. How are you experiencing the results and recommendations of the national discipline assessment of medical and health research of 1998?

Very useful		**Neutral**		**Not useful**
1	2	3	4	5

Question 47b. How are you experiencing the results and recommendations of research school evaluations in general?

Very useful		**Neutral**		**Not useful**
1	2	3	4	5

Question 47c. Please indicate whether the results and recommendations of the national discipline assessment of medical and health research of 1998 are incorporated?

Very intensively		**Neutral**		**Not intensively**
1	2	3	4	5

Question 47d. Please indicate whether the results and recommendations of the research school evaluations are incorporated in general?

Very intensively		**Neutral**		**Not intensively**
1	2	3	4	5

Question 48. What is your opinion about the following general statements about the national discipline assessment of medical and health research of 1998

	Dis-agree fully	**Dis-agree partly**	**Neither agree nor disagree**	**Agree partly**	**Agree fully**
1. The results and recommendations are relevant because research is judged on actual value	❑	❑	❑	❑	❑
2. The results and recommendations are not relevant because research is not assessed at the group level	❑	❑	❑	❑	❑
3. The results and recommendations are not relevant because the research assessed is not up to date	❑	❑	❑	❑	❑
4. The results and recommendations are not relevant because in my opinion inappropriate criteria are used	❑	❑	❑	❑	❑
5. I attach higher value to results of internal evaluations	❑	❑	❑	❑	❑
6. I support the replacement of national discipline assessment of medical and health research by a new external research evaluation protocol, consisting of self-evaluation and periodical external evaluations	❑	❑	❑	❑	❑

Question 49. Please indicate if you agree with the statement: 'The evaluation of the social quality of research will stimulate researchers not only to perform well from a scientific but also from societal point of view'.

1 ❑ Yes
2 ❑ No
3 ❑ Don't know

Notes:
[1] Questions 17 and 18 are after Colquitt (2001)

ANNEX 5

Correlation matrix of research performance measures

	SCI publications		DAG score 1998		External funding		Proposals NWO		Proposals CMHR		CMHR grants		NWO rating		NWO activities
SCI publications															
DAG score 1998	Pre-clin	0.10													
	Clinical	0.26													
External funding	Pre-clin	0.12	Pre-clin	**0.33***											
	Clinical	-0.03	Clinical	0.22											
Proposals NWO	Pre-clin	**0.29***	Pre-clin	0.23	Pre-clin	**0.24***									
	Clinical	0.21	Clinical	**0.32***	Clinical	0.26									
Proposals CMHR	Pre-clin	0.23	Pre-clin	0.18	Pre-clin	0.20	Pre-clin	**0.82*****							
	Clinical	0.06	Clinical	0.21	Clinical	0.18	Clinical	**0.78*****							
CMHR grants	Pre-clin	0.22	Pre-clin	0.23	Pre-clin	0.19	Pre-clin	**0.71*****	Pre-clin	**0.91*****					
	Clinical	0.16	Clinical	**0.35***	Clinical	0.08	Clinical	**0.50*****	Clinical	**0.75*****					
NWO rating	Pre-clin	0.25*	Pre-clin	**0.34***	Pre-clin	0.20	Pre-clin	**0.52*****	Pre-clin	**0.48*****	Pre-clin	**0.45*****			
	Clinical	0.22	Clinical	0.26	Clinical	**0.31***	Clinical	**0.58*****	Clinical	**0.47*****	Clinical	**0.29***			
NWO activities	Pre-clin	**0.28***	Pre-clin	0.23	Pre-clin	0.24	Pre-clin	**0.97*****	Pre-clin	**0.93*****	Pre-clin	**0.85*****	Pre-clin	**0.53*****	
	Clinical	0.18	Clinical	**0.32***	Clinical	0.23	Clinical	**0.95*****	Clinical	**0.92*****	Clinical	**0.73*****	Clinical	**0.55*****	

$^{*}p < 05$; $^{**}p <.01$; $^{***}p <0.001$

ANNEX 6

Correlation matrix of contingencies with research performance

		SCI publications	DAG score	External funding	Proposals NWO	Proposals CMHR	CMHR grants	NWO rating	NWO activities	NWO activities dummy
Size	Pre-clin	0.06	0.03	0.22	**0.46*** **	**0.46*** **	**0.38*** **	0.18	**0.48*** **	
Research staff	Clinical	**0.50*** **	**0.36***	0.13	0.05	0.00	0.15	0.11	0.06	0.11
Work setting	Pre-clin	**-0.27***	**0.35***	**0.36****	-0.09	-0.05	-0.01	0.04	-0.07	
	Clinical	-0.12	0.17	0.23	0.07	0.01	-0.01	-0.09	0.03	0.07
Time allocation										
Research	Pre-clin	0.02	**0.67*** **	0.18	0.13	0.22	**0.27***	**0.30***	0.17	
	Clinical	**0.42*** **	0.08	-0.10	0.14	0.07	0.09	0.06	0.13	0.03
Education	Pre-clin	0.05	**-0.31***	-0.10	-0.11	-0.17	-0.16	-0.07	-0.17	
	Clinical	0.08	-0.07	-0.19	-0.22	-0.10	-0.02	-0.03	-0.16	-0.01
Patient care	Pre-clin	0.05	**-0.63*** **	**-0.47*** **	**-0.26***	-0.22	**-0.24***	**-0.33****	**-0.28***	
	Clinical	0.12	-0.24	-0.05	-0.05	0.03	0.00	0.04	-0.02	0.02
Supervision of PhD students	Pre-clin	**0.31***	**0.51*** **	**0.27***	0.18	0.19	0.22	**0.29***	0.19	
	Clinical	0.12	0.11	0.07	0.16	-0.07	0.04	-0.02	0.09	0.05
Age										
Research experience	Pre-clin	0.12	0.08	-0.14	-0.13	-0.23	**-0.27***	-0.15	-0.18	
	Clinical	0.19	-0.08	0.04	-0.08	-0.06	-0.07	0.03	-0.08	0.02
	Pre-clin	-0.05	-0.26	-0.21	-0.20	-0.20	-0.23	**-0.31****	-0.20	
Management experience	Clinical	-0.19	-0.20	-0.07	-0.17	-0.11	-0.17	-0.07	-0.17	-0.20
	Pre-clin	**0.24***	0.18	0.03	0.17	0.07	0.03	0.13	0.13	
Foreign Experience	Clinical	**0.31***	-0.02	0.07	0.15	0.12	0.22	-0.09	0.18	-0.02
	Pre-clin	0.20	0.02	-0.02	-0.09	-0.23	**-0.27***	0.00	-0.17	
Supervising experience	Clinical	0.18	0.15	**0.27***	0.10	0.15	**0.29***	0.18	0.17	0.18

$^{*}p < 05$; $^{**}p < .01$; $^{***}p < 0.001$

ANNEX 7

Correlation matrix of internal control with research performance

		SCI publications	DAG score 1998	External funding	Proposals NWO	Proposals CMHR	CMHR grants	NWO rating	NWO activities	NWO activities dummy
Control of personal resources	Pre-clin	0.18	-0.02	0.15	0.18	**0.26***	0.25	0.23	0.22	
	Clinical	-0.09	0.01	0.03	0.02	0.01	0.21	-0.01	0.04	-0.01
Research commitment	Pre-clin	**0.26***	**0.43****	0.14	0.10	0.11	0.16	**0.27***	0.11	
	Clinical	0.10	0.00	0.07	0.09	0.09	0.01	-0.02	0.08	0.09
Time spent on internal managerial control	Pre-clin	0.12	0.25	**0.26***	0.21	0.20	0.23	**0.37****	**0.24***	
	Clinical	0.05	-0.04	0.04	0.23	0.15	**0.27***	0.17	0.24	0.16
Communication	Pre-clin	**0.24***	**0.44****	-0.01	0.04	0.03	0.13	0.10	0.04	
Research meetings	Clinical	**0.33***	0.00	0.12	0.13	0.20	0.17	0.22	0.17	0.20
	Pre-clin	-0.02	-0.09	-0.06	0.09	0.04	-0.05	-0.21	0.07	
Direct communication	Clinical	-0.06	-0.13	-0.06	0.00	0.06	0.10	0.09	0.04	0.03
Rewards	Pre-clin	0.05	**0.31***	**0.32***	0.17	0.05	0.07	0.23	0.11	
Development of research	Clinical	**0.27***	0.20	0.14	0.23	0.09	0.08	0.15	0.18	0.18
skills	Pre-clin	0.06	0.01	0.22	**0.28***	**0.26***	0.17	0.15	**0.28***	
	Clinical	**-0.31***	-0.14	-0.13	-0.15	-0.11	-0.16	-0.18	-0.16	-0.23
Flexibility	Pre-clin	**0.31***	0.12	**0.35****	**0.29***	0.19	0.18	0.14	**0.27***	
	Clinical	**0.29***	0.24	**0.42****	0.10	0.00	-0.09	0.17	0.04	**0.31***
Special commendations	Pre-clin	0.02	0.03	-0.07	0.23	0.18	0.15	-0.04	0.23	
	Clinical	**0.32***	-0.02	0.19	0.08	-0.07	0.04	0.23	0.04	0.15
Financial bonus system										
Research policy planning	Pre-clin	0.09	-0.07	0.14	**0.25***	0.20	0.15	**0.32****	**0.24***	
	Clinical	-0.03	0.10	-0.03	0.08	0.19	0.19	0.22	0.14	0.18
Job evaluation	Pre-clin	-0.01	0.06	0.03	-0.23	-0.18	-0.11	-0.11	-0.21	
	Clinical	-0.05		-0.06	0.05	0.09	0.06	-0.11	0.07	-0.10

Pre-evaluation research proposals	Pre-clin	**0.24***	-0.01	0.08	0.15	0.10	0.00	0.07	0.10	
Group/Department	Clinical	0.06	-0.11	0.02	0.07	0.10	0.03	0.12	0.08	-0.05
	Pre-clin	-0.01	-0.23	0.03	-0.23	-0.16	**-0.27***	-0.16	**-0.25***	
Medical Faculty	Clinical	0.21	0.22	-0.01	0.02	0.09	0.10	0.16	0.06	0.08
	Pre-clin	-0.10	-0.05	**0.41*****	-0.05	0.11	0.07	-0.04	-0.01	
University / Non-university institute	Clinic	0.03	-0.20	0.01	0.09	0.13	0.05	0.11	0.11	0.07
Research output evaluation	Pre-clin	0.02	-0.25	-0.07	-0.09	-0.11	-0.15	-0.22	-0.10	
Group/Department	Clinical	0.08	0.10	**0.28***	0.07	0.10	0.03	0.12	0.08	0.08
	Pre-clin	0.05	-0.25	0.06	0.02	-0.05	-0.10	-0.17	-0.01	
Medical Faculty	Clinical	0.04	0.22	**0.34***	0.23	0.23	0.13	**0.25***	0.23	**0.28***
	Pre-clin	0.07	0.16	0.14	0.02	0.00	-0.08	-0.10	0.01	
University/Non-university institute	Clinical	-0.03	-0.06	0.17	0.05	0.03	0.12	0.25	0.07	**0.27***
Effectiveness of pre-evaluations	Pre-clin	0.02	-0.08	0.06	-0.04	-0.07	0.02	-0.05	-0.05	
	Clinical	0.09	0.06	0.18	**0.29***	0.21	0.16	-0.03	**0.27***	0.10
Importance of internal research evaluations	Pre-clin	0.11	-0.16	-0.04	0.00	0.00	-0.05	0.03	0.01	
	Clinical	-0.20	0.16	0.11	-0.08	-0.14	-0.21	-0.13	-0.13	-0.07

$^{*}p < 05$; $^{**}p < .01$; $^{***}p < 0.001$

ANNEX 8

Correlation matrix of external control with research performance

		SCI publications	DAG score 1998	External funding	Proposals NWO	Proposals CMHR	CMHR grants	NWO rating	NWO activities	NWO activities dummy
Time spent on external research activities	Pre-clin	0.16	0.02	0.03	**0.28***	0.16	0.18	0.12	**0.25***	
	Clinical	**0.47****	0.12	0.19	0.13	-0.07	-0.02	0.10	0.06	0.23
Time spent on external managerial control	Pre-clin	0.17	0.27	**0.37****	**0.32****	0.19	0.10	0.22	**0.31***	
	Clinical	-0.02	**0.36***	**0.40****	0.11	0.09	0.10	0.02	0.12	0.18
Importance of external research evaluations										
DAG evaluations 1998	Pre-clin	-0.14	-0.21	-0.11	-0.13	-0.07	0.06	-0.04	-0.05	
	Clinical	-0.12	-0.25	-0.02	-0.13	-0.04	-0.14	-0.21	-0.12	-0.18
Research school evaluation	Pre-clin	-0.14	-0.01	0.10	0.12	0.00	0.08	0.25*	0.09	
	Clinical	0.08	-0.15	0.05	0.11	0.17	0.24	-0.19	0.16	-0.04
Importance of collaboration										
Research groups	Pre-clin	0.10	-0.08	-0.06	-0.02	0.01	0.08	0.09	0.00	
	Clinical	-0.08	-0.01	0.07	-0.23	-0.10	-0.01	-0.17	-0.17	-0.14
Ministries and semi-governmental organisations	Pre-clin	0.02	**-0.31***	-0.10	-0.13	-0.07	-0.08	-0.22	-0.10	
	Clinical	0.08	**0.33***	-0.10	-0.04	-0.05	-0.03	0.10	-0.05	0.09
Pharmaceutical industries & firms	Pre-clin	0.01	-0.13	-0.05	-0.07	0.02	0.07	0.01	-0.03	
	Clinical	0.08	**0.39***	**0.29***	-0.08	-0.11	-0.12	-0.04	-0.11	0.06
Clinic	Pre-clin	0.05	**-0.35***	**-0.27***	-0.19	-0.08	-0.03	-0.21	-0.15	
	Clinical	0.10	-0.21	-0.13	-0.13	-0.08	-0.13	**-0.26***	-0.13	**-0.27***
Family doctor	Pre-clin	0.03	**-0.30***	-0.21	-0.12	-0.05	-0.05	**-0.26***	-0.08	
	Clinical	0.06	0.11	0.15	-0.02	0.00	-0.02	0.18	0.01	0.17

$^{*}p < 05$; $^{**}p <.01$; $^{***}p <0.001$

ANNEX 9

Correlation matrix of contingencies with internal control

		Size	Work setting	Time to research	Time to education	Time to patient care	Time to supervision	Research experience	Management experience	Foreign experience	Supervising experience
Control of personal resources	Pre-clin	0.23	0.14	0.15	-0.19	-0.05	0.23	-0.18	-0.11	**0.26***	-0.08
	Clinical	0.03	0.12	0.23	-0.06	0.05	0.20	0.08	**0.43***	-0.02	0.08
Research commitment	Pre-clin	-0.22	**0.28***	**0.45*****	-0.10	**-0.33****	**0.47*****	0.08	-0.07	0.19	0.17
	Clinical	-0.17	0.19	**0.29***	-0.11	-0.09	**0.37****	0.14	-0.24	0.13	0.10
Time spent on internal managerial control	Pre-clim	0.18	0.14	0.19	-0.17	**-0.37*****	**0.43*****	0.00	-0.01	0.08	0.03
	Clinical	0.22	-0.01	**0.35****	-0.02	**-0.35****	**0.40****	-0.14	-0.11	0.08	000
Communication	Pre-clin	-0.16	0.13	**0.29***	-0.07	-0.08	**0.27***	-0.14	**-0.29***	-0.12	-0.03
Research meetings	Clinical	0.15	-0.06	0.09	0.12	0.06	0.12	-0.01	-0.14	0.08	0.07
	Pre-clin	0.11	-0.14	**-0.40*****	-0.13	0.14	-0.14	0.05	0.21	-0.15	-0.10
Direct communication	Clinical	-0.01	-0.18	-0.12	0.03	0.16	-0.06	0.08	0.26	0.05	-0.20
Rewards											
Development skills	Pre-clin	0.05	0.07	0.13	0.02	**-0.29***	0.05	-0.07	-0.18	**0.28***	0.01
	Clinical	0.22	0.06	0.14	0.14	-0.12	0.18	0.12	-0.04	**0.27***	0.13
Flexibility	Pre-clin	0.13	0.04	0.09	0.10	-0.03	-0.12	**-0.24***	-0.06	**0.24***	-0.16
	Clinical	-0.19	-0.04	0.17	0.12	**-0.32***	0.07	-0.06	0.05	0.08	-0.09
Special commendations	Pre-clin	0.14	0.11	0.01	-0.06	-0.01	0.03	-0.07	0.07	0.07	0.01
	Clinical	**0.37****	0.07	-0.04	0.03	0.18	0.03	0.17	-0.14	0.22	0.11
Financial bonus system	Pre-clin	**0.34****	**-0.29***	-0.06	**0.24***	0.15	-0.14	-0.16	0.01	0.05	**-0.33****
	Clinical	**0.28***	-0.03	-0.02	-0.01	0.05	-0.03	0.05	-0.01	0.20	0.07
Research policy planning	Pre-clin	**0.27***	**-0.28***	-0.04	0.06	-0.05	-0.17	-0.03	-0.14	-0.04	-0.03
	Clinical	0.11	-0.12	0.11	0.23	-0.09	-0.01	0.04	**0.27***	0.14	-0.13
Job evaluation	Pre-clin	0.00	0.03	0.00	-0.08	0.04	-0.05	0.04	0.02	-0.04	0.09
	Clinical	0.10	0.04	-0.11	-0.07	-0.12	0.04	-0.23	0.10	-0.12	0.03

Pre-evaluation of research proposals	Pre-clin	0.15	-0.06	0.12	-0.01	0.14	0.08	0.07	-0.14	0.05	0.17
Group/Department	Clinical	-0.15	0.09	0.20	-0.07	-0.17	-0.07	0.08	0.16	0.03	0.15
	Pre-clin	-0.03	-0.03	-0.04	0.03	0.19	-0.12	-0.11	-0.01	-0.02	0.02
Medical Faculty	Clinical	0.04	0.07	-0.10	-0.09	0.17	-0.07	0.17	0.03	-0.02	0.09
	Pre-clin	0.18	**0.39****	0.21	-0.22	0.03	0.06	-0.16	-0.09	-0.12	-0.07
University / Non-university institute	Clinic	-0.18	0.07	0.10	-0.03	0.16	-0.07	0.08	0.16	-0.05	-0.04
Research output evaluation	Pre-clin	-0.08	-0.05	-0.08	0.23	0.22	-0.05	0.01	-0.02	0.04	0.05
Group/Department	Clinical	0.11	0.11	0.09	0.01	-0.06	-0.14	0.22	0.00	0.17	**0.39****
	Pre-clin	0.08	-0.17	-0.20	0.12	0.18	-0.22	-0.09	0.00	0.07	-0.10
Medical Faculty	Clinical	0.06	-0.08	0.13	**0.35****	-0.06	-0.08	0.09	-0.01	0.00	0.23
	Pre-clin	0.18	0.07	0.02	-0.13	0.21	-0.09	0.08	-0.01	-0.09	-0.10
University/Non-university institute	Clinical	-0.03	0.02	0.15	0.01	0.03	0.17	-0.11	0.04	0.14	-0.25
Effectiveness of pre-evaluations	Pre-clin	-0.14	**0.27***	0.09	-0.16	0.03	0.05	0.17	-0.05	0.19	0.21
	Clinical	-0.18	0.16	**0.26***	-0.06	-0.07	0.13	-0.08	0.02	0.17	0.21
Importance of internal research evaluations	Pre-clin	0.03	0.03	-0.18	-0.17	0.10	**-0.25***	0.00	0.05	0.19	-0.01
	Clinical	0.08	0.11	-0.05	0.01	0.01	0.12	0.02	-0.09	0.07	-0.07

$^{*}p < 05$; $^{**}p <.01$; $^{***}p <0.001$

ANNEX 10

Correlation matrix of contingencies with external control

		Size	Work setting	Time to research	Time to education	Time to patient care	Time to supervision	Research experience	Management experience	Foreign experience	Supervising experience
Time spent on external research activities	Pre-clin	0.10	-0.18	0.14	0.13	0.19	-0.06	-0.02	-0.13	0.08	-0.01
	Clinical	**0.29***	-0.07	**0.33***	0.09	-0.05	**0.27***	0.13	-0.21	**0.26***	0.01
Time spent on external managerial control	Pre-clin	**0.25***	0.08	0.07	-0.13	**-0.29***	0.12	0.10	-0.19	-0.01	-0.01
	Clinical	**0.32***	0.17	-0.01	-0.01	-0.25	0.01	0.24	-0.02	0.17	0.16
Importance of external research evaluations											
DAG evaluation 1998	Pre-clin	0.04	-0.14	-0.19	0.17	0.03	0.12	-0.08	0.03	-0.24	0.05
	Clinical	-0.22	-0.03	-0.16	0.03	0.06	-0.04	0.05	-0.07	0.01	0.12
Research school evaluations	Pre-clin	-0.09	0.14	0.13	-0.04	-0.11	0.13	-0.23	-0.16	0.18	-0.07
	Clinical	-0.07	0.12	-0.06	-0.24	0.02	0.16	0.05	0.10	0.08	0.31
Importance of collaboration											
Research groups	Pre-clin	-0.12	-0.14	-0.13	0.00	0.07	-0.15	-0.12	-0.03	-0.11	-0.10
	Clinical	0.09	0.00	-0.12	-0.04	0.04	-0.03	0.02	0.03	**0.28***	0.06
Ministries and semi-governmental organisations	Pre-clin	0.15	**-0.27***	-0.22	0.05	**0.41*****	**-0.31***	-0.08	0.22	-0.15	**-0.35****
	Clinical	0.15	0.05	-0.01	0.18	-0.06	0.19	-0.19	0.10	-0.06	-0.22
Pharmaceutical industries & firms	Pre-clin	0.07	-0.06	0.09	0.23	0.23	-0.19	**-0.26***	-0.04	-0.16	**-0.45*****
	Clinical	0.16	0.14	0.02	-0.05	0.09	-0.12	-0.02	0.08	**0.37****	0.10
Clinic	Pre-clin	0.06	-0.23	**-0.25***	**-0.25***	**0.34****	-0.23	-0.04	0.16	**-0.31***	-0.15
	Clinical	-0.22	0.00	**-0.26***	0.03	**0.44*****	-0.25	0.08	0.22	0.24	-0.15
Family doctor	Pre-clin	0.08	**-0.29***	**-0.26***	**0.31***	**0.34****	**-0.27***	0.06	**0.42****	-0.08	**-0.25***
	Clinical	0.03	0.04	-0.25	0.24	-0.09	0.19	-0.24	0.07	-0.22	-0.08

$^{*}p < 05$; $^{**}p < .01$; $^{***}p < 0.001$

DANKWOORD

Allereerst bedank ik alle respondenten (wetenschappers en beleidsmedewerkers) die bereid waren mee te werken en hun inzichten en kennis met mij wilden delen aan dit onderzoek. Zonder jullie geen empirisch materiaal en dus geen proefschrift.

Mijn complimenten aan mijn (co)promotoren. Peter, jouw enthousiasme zorgde ervoor dat ik het aandurfde om te beginnen aan een promotie traject. Bedankt dat je me deze mogelijkheid hebt gegeven. Je bent een echte ideeën generator en hebt de gave om opmerkingen te kunnen plaatsen die de kern van de zaak raken (als ik weer eens een zijweg was ingeslagen). Ik heb de afgelopen jaren enorm veel van je geleerd. Jij hebt mij veel ruimte en vrijheid gelaten om binnen het project mijn eigen weg te gaan. Ook stimuleerde je me om verder dan alleen het onderzoeksproject te kijken. Veel dank daarvoor! Eduard, bedankt voor je enthousiasme, betrokkenheid, steun en vertrouwen in mij. Ondanks je overvolle agenda wist je toch altijd weer tijd vrij te maken voor efficiënte besprekingen, eerst in de stationsrestauratie van Leiden (onder het genot van een koffie verkeerd), later in je eigen kamer in het LUMC. De sfeer tijdens de besprekingen was altijd ontspannen. Jouw streven naar helderheid en duidelijkheid hielp mij 'to-the-point' te blijven. Ik hoop dat je trots kan zijn op het eindproduct! Dick, zonder jou zou er heel wat minder kwantitatief materiaal in het proefschrift zijn gekomen. Bedankt dat jij, halverwege het promotietraject, in de begeleidingscommissie wilde stappen. Je hebt me rondgeleid in de SPSS wereld en ik kon altijd bij je terecht als ik weer eens was vastgelopen in de data. Ik waardeer je wetenschappelijke nauwkeurigheid en precisie, ook al maakte dit mij wel eens ongeduldig. Jouw kritische blik op de concepten van hoofdstukken heeft zeker geresulteerd in kwaliteitsverbetering! Kortom, een sociale wetenschapper, een psycholoog en een medicus als begeleiders -op het eerste gezicht een lastige combinatie- jullie inhoudelijk inbreng was divers, maar juist omdat jullie elkaar aanvulden vormden we samen een uitstekend team!

Het onderzoek zoals beschreven in dit proefschrift was niet mogelijk geweest zonder verdere financiële steun van NWO. Hierbij wil ik ook ZonMw bedanken voor de unieke kans die ze me boden om praktijkervaring op te doen. Ook bij het kwantitatieve gedeelte in het promotie onderzoek kon ik rekenen op jullie hulp. Daarnaast vond ik het heel leuk om de uiteindelijke resultaten van het kwantitatieve gedeelte te hebben mogen presenteren tijdens het zogenaamde 'broodje NWO'. Tenslotte Team 1 bedankt voor jullie interesse in mijn onderzoek en de interessante discussies!

De leden van de begeleidingscommissie bedank ik voor de tijd die ze gestoken hebben in het lezen van conceptstukken en het geven van kritisch commentaar. Helen Richardson wil ik bedanken voor de Engelse correcties.

De eerste twee jaar is het onderzoek uitgevoerd in de MAWIN groep. Hierbij wil ik mijn oud-MAWIN collega's bedanken voor de interesse in mijn onderzoek, de fijne samenwerking in onderwijsprojecten, en de gezellige sfeer. Veel dank gaat uit naar de afdeling Bestuur en Organisatie die mij halverwege het promotietraject de mogelijkheid gaf om mijn project verder uit te voeren en af te ronden. De overstap van Exacte Wetenschappen naar de Sociale Wetenschappen was niet altijd makkelijk, we spraken een andere taal. Toch heb ik het gevoel een van jullie te zijn geworden, ik wil dan ook mijn Strategie en Organisatie collega's bedanken voor jullie interesse in mijn onderzoek en de inspirerende discussies. Ook mede promovendi van de afdelingen Bestuur en Organisatie, Communicatiewetenschap en Politicologie: bedankt voor de discussies over het wel-en-wee van het AIO-schap en het sociale netwerk. Ik denk met veel plezier terug aan de gezellige etentjes in de Bodega (de gamba's!) en de summerschools in Essex.

Scriptiestudenten Anna, Maaike en Swi-Fan: het was erg leuk en inspirerend om met jullie te werken, heel veel succes in jullie loopbanen! Ook Carlo wil ik bedanken voor invoeren van geretourneerde vragenlijsten in SPSS, jij hebt me veel werk uit handen genomen!

Ook een dank je wel aan de dames van het AIO clubje. Ik heb veel gehad aan onze besprekingen en discussies. Bedankt voor het begrip, de steun, de interesse en het commentaar op mijn concepten.

Pap en mam: bedankt voor de mogelijkheid en vrijheid die jullie me gegeven hebben om me te ontwikkelen en ontplooien. Mijn vriendinnen, vrienden en (schoon)familieleden bedank ik voor de fijne open gesprekken, gezellige etentjes en sportieve uitjes. Jullie steun en peptalks hebben mij goed gedaan!

Mijn laatste woord van dank gaat uit naar Jim. Jij bent de rots in de branding, mijn uitlaatklep, mijn luisterend oor. Bij jou kan ik mijn hart uit storten, jij blijft altijd rustig en kalm. Jouw geloof in mij heeft me enorm gestimuleerd dit proefschrift af te ronden.

Utrecht, Oktober 2006

OVER DE AUTEUR

Inge van der Weijden werd op 22 september 1976 geboren te Heemstede. In 1994 behaalde zij haar VWO diploma en begon aan de Vrije Universiteit met de studie Medische Biologie. Ze koos voor de afstudeerrichting beleid en management. In 1998 heeft ze als afstudeeropdracht (onder leiding van voormalig directeur prof. Joost Ruitenberg) bij het Centraal Laboratorium van de Bloedtransfusiedienst (tegenwoordig Sanquin) het functioneren van het stakeholdersmodel geëvalueerd. Met de analyse van dit managementsysteem voor medisch wetenschappelijk onderzoek is de fascinatie voor onderzoeksmanagement en evaluatie ontstaan. In 1999 studeerde zij af en begon als junior onderzoeker bij de afdeling MAWIN (faculteit Exacte Wetenschappen) aan de Vrije Universiteit met een verkenning van het promotie onderwerp. In 2000 ontving zij een subsidie van NWO om gedurende een periode van drie jaar empirisch onderzoek te verrichten. In 2001 heeft zij 5 maanden praktijk ervaring opgedaan bij NWO Medische Wetenschappen (tegenwoordig CMHR). Ze werkte als programmasecretaris divisie orgaansystemen en heeft mogen ervaren hoe het beoordelingsproces van onderzoeksvoorstellen in de praktijk plaatsvond. Eind 2001 maakte ze de overstap naar de afdeling Bestuur en Organisatie (faculteit Sociale Wetenschappen) om in nabijheid van haar co-promotoren en in een inspirerend onderzoeksklimaat verder aan het promotie onderzoek te werken. Daarnaast was zij eindredacteur van het tijdschrift Wetenschap Technologie en Samenleving (2000-2002) en behartigde ze op facultair niveau de belangen van promovendi (2002). Na afloop van haar aanstelling in 2004 heeft zij een jaar als programmasecretaris in het ZonMw Open Programma gewerkt. In 2005 was ze werkzaam als research fellow bij de afdeling Bestuur en Organisatie en in het voorjaar van 2006 als docent bij de afdeling Methoden en Technieken aan de Vrije Universiteit. Vanaf oktober 2006 werkt ze als onderzoeker bij het Rathenau Instituut aan het Science System Assessment programma.

DISSERTATION SERIES FACULTY OF SOCIAL SCIENVES, VRIJE UNIVERSITEIT AMSTERDAM

Does Succes Breed Succes? Effects of news and advertising on corporate reputation, May-May Meijer, December 2004, Aksant Academic Publishers: Amsterdam, isbn 90 5260 145 3.

The Polder Model: From Disease to Miracle? Dutch Neo-corporatism 1965-2000, Jaap Woldendorp, June 2005, Amsterdam, isbn 90 77472 05 3.

Interviewer and Respondent. Interaction in Survey Interviews, Yfke Ongena, December 2005, Amsterdam, isbn 90 9020070 3.

Integriteitsbeleid van de Nederlandse politie, Terry Lamboo, december 2005, Eburon: Delft, isbn 90 5972 088 L.

Politics of Sovereignty, Tanja E. Aalberts, June 2006, Amsterdam, isbn 10: 90 9020656 6 / 13: 978 90 9020656 1.

In Pursuit of Comfort. The Transnationalisation Process of Malaysian Chinese Small and Medium Enterprises, Esther Zwart, November 2006, Amsterdam, isbn 90 8659024 1.

Intergenerational relationships of older adults. Family structure, contact and norm, Suzan van der Pas, November 2006, Amsterdam, isbn-10: 90-9021075-X, isbn-13: 978-90-9021075-9.